UNIVERSITY OF CALIFORNIA
RIVERSIDE

Metabolite Assignment and Profiling of Environmental Stressors in Earthworms (*Eisenia fetida*), Coelomic Fluid, and Coelomocytes

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Corey Michael Griffith

March 2019

Dissertation Committee:
Dr. Cynthia K. Larive, Chairperson
Dr. Jay Gan
Dr. Wenwan Zhong

The Dissertation of Corey Michael Griffith is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Graduate school was a challenging, but gratifying adventure. I would like to deeply thank my mentor, Professor Cindy Larive, for her support, insight, and dedication. I am very appreciative of the many opportunities I have received, including the opportunity to start a brand-new project upon joining her lab. The guidance and flexibility gave me the space to explore, learn, and grow. I am very proud of our work, and the breath of science it opened me up to. I would like to thank my labmates: Drs. Meredith Dinges, Melissa Morgan, and Andrew Green, in addition to our postdoctoral fellow Dr. Caroline Mathon, and visiting scholars Drs. Luzineide Tinoco and Alvicler (Calvin) Magalhaes. It was a pleasure to work with each of you, and I cherish the advice, support, and friendship we share. I would also like to thank all my other science mentors over the years, including Drs. Jim Seiber, Ron Tjeerdema, Caitlin Reering, and Rebecca Mulligan at UC Davis and Dr. John Beck, Nausheena Baig, and Wai Gee at the U.S. Department of Agriculture.

I would like to acknowledge of the Environmental Toxicology Graduate Program and UC Riverside, and the financial support from the National Institute of Environmental Health Sciences T32 Training Grant (T32 ES018827), Graduate Research Mentorship Program Fellowship, Dissertation Year Program Fellowship, GSA and Fukuto travel grants, and UCR Metabolomics Core Facility seed grant. I would like to also acknowledge support from the American Chemical Society (ACS) and Society of Environmental Toxicology and Chemistry (SETAC) for travel awards, and ACS Environmental Chemistry Division for serving as a Graduate Service Fellow.

I could not have been successful in graduate school without the love and support of my family and friends. My parents, Gail and Pat, my sisters, Kaylee and Sherry, and my niece, Skyler are always there for me, and I love and value each of you. There is something special with the friendships and bonds made in graduate school. It can be a tough environment, and the people you see every day make all the difference. I particularly would like to acknowledge Sam Hinman, Preston and Nicole Williams, Yuxiang Cui, Kristy McKeating, and Marissa Giroux in addition to the many others in and out of graduate school that have made a huge impact on my life. I love and appreciate all of you and thank you for always being there.

Acknowledgements of Copyright

ABSTRACT OF THE DISSERTATION

Metabolite Assignment and Profiling of Environmental Stressor in Earthworms (*Eisenia fetida*), Coelomic Fluid, and Coelomocytes

by

Corey Michael Griffith

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, March 2019
Dr. Cynthia K. Larive, Chairperson

Earthworms (*Eisenia fetida*) are abundant and vital members of the soil environment and monitoring their metabolism may be a useful indicator of soil health and ecotoxicity. This dissertation aimed to expand the analysis of nonlethal and noninvasive earthworm metabolite pools to probe their usefulness in environmental monitoring. Coelomic fluid (CF) is a biofluid that fills the body cavity of the worm and contains free-moving liver-like and immune cells called coelomocytes (CC). ^{1}H NMR, GC-MS, and LC-MS were used to identify metabolites in earthworm, CF, and CC extracts. Fifty-four metabolites were detected in earthworm extracts, 47 in CF, and 41 in CC using ^{1}H NMR. GC-MS was only employed with CF extracts, where 44 metabolites were detected. Targeted LC-MS analyses detected 97 metabolites in earthworm extracts, 82 in CF, and 67 in CC. Significantly, assignment of the earthworm metabolomes led to the identification of a new metabolite: (–)-*β*-L-malyl-L-glutamate. Malylglutamate was detected in several invertebrate species at concentrations in the ng/mg – μm/mg range, and it was elucidated as a chelator and

potential store for malate and glutamate. This dissertation also sought to explore the response of earthworm metabolism to environmental stressors. CF of earthworms exposed to six chloroacetanilide herbicides (acetochlor, alachlor, butachlor, metolachlor, S-metolachlor, and propachlor) were collected and analyzed using ^{1}H NMR and GC-MS. Perturbations in lipid metabolism and β-oxidation were observed, suggesting that chloroacetanilide herbicides affected earthworms in a manner similar to the herbicidal mode of action. Exposure to chlorothalonil was used to compare the impact of metabolites in earthworm, CF, and CC extracts using ^{1}H NMR and LC-MS. CF extracts were the most sensitive matrix to detect the effects of chlorothalonil exposure, where increased glutamine levels was the only biomarker detected at both doses. Chemometrics revealed *N*-acetylserine and ophthalmic acid as strong biomarkers in the high dose group of CF extracts, which may indicate increased oxidative stress. ADP ribose was the only metabolite consistently affected and its increase could be a response to chlorothalonil-induced DNA damage. This work supports metabolic profiling in earthworm and CF extracts to determine which matrix is most sensitive for detecting environmental stress.

Table of Contents

Metabolite Assignment and Profiling of Environmental Stressors in Earthworms (*Eisenia fetida*), Coelomic Fluid, and Coelomocytes

List of Figures

List of Tables

CHAPTER ONE

Introduction

Based in part on papers published in the *Journal of Proteome Research*

J. Proteome Res. 2017, *16* (9), 3407-3417

J. Proteome Res. 2018, *17* (8), 2611-2622

This dissertation research aims to develop new methods for using metabolomics to identify the non-targeted, sub-lethal effects of toxicants in the soil environment, using the earthworm species *Eisenia fetida*. Improvements in methods for monitoring environmental health and advancing our understanding of environmental effects depends on advances in technologies for chemical analysis and ecotoxicity assessment. Current limits of detection strain the true characterization of chemicals in the environment, some of which may hold significant toxicological implications at their low, unquantifiable levels or interact as components of a complex mixture through synergistic effects. Thus, alternative methods are needed to identify contaminants and evaluate ecotoxicity. One such approach is environmental metabolomics, which seeks to understand how organisms respond to toxicants by analyzing changes in metabolite levels within cells, tissues, or biofluids.[1-2]

Earthworms are abundant and vital members of the soil environment. They are sensitive to contaminants and monitoring their metabolism may be a useful indicator of soil health and ecotoxicity. Herein, this dissertation aimed to study complementary, nonlethal, and noninvasive metabolite pools to probe their usefulness in environmental monitoring. These studies first required identifying earthworm metabolites to allow

interpretation of biological perturbations, which led to the elucidation of a new, abundant earthworm metabolite: malylglutamate. A significant portion of this dissertation also focuses on elucidating the structure of malylglutamate and screening for its presence in invertebrates. Nuclear magnetic resonance spectroscopy (NMR), gas chromatography-mass spectrometry (GC-MS), and liquid chromatography-mass spectrometry (LC-MS) are used orthogonally to understand earthworm metabolism and its response to environmental stressors through the following objectives:

Objective 1: Identify and assign ^{1}H NMR resonances of metabolites in *E. fetida* coelomic fluid, coelomocyte, and whole-earthworms. (Chapter 2)

Objective 2: Elucidate the structure and absolute configuration of the major earthworm metabolite, malylglutamate. (Chapter 3)

Objective 4: Explore the function of malylglutamate and screen for its presence in closely-related invertebrates. (Chapter 4)

Objective 5: Profile the metabolic impacts of chloroacetanilide herbicide exposure in earthworm coelomic fluid and identify potential structure-activity relationships using ^{1}H NMR and GC-MS. (Chapter 5)

Objective 6: Compare the metabolic impacts and identify biomarkers of chlorothalonil exposure in coelomic fluid, coelomocytes, and whole-earthworms using ^{1}H NMR and targeted LC-MS. (Chapter 6)

1.1. Environmental Metabolomics and Ecotoxicology

Assessing ecotoxicity is an arduous task, with an understanding of the effects of a contaminant often delayed until adverse effects are seen in species higher on the trophic level. There is a need to identify environmental insults prior to ecosystem-wide problems. Indicator species or bioindicators are used monitor the health and integrity of an ecosystem, analogous to the canary in the coal mine.[3] A good indicator species is representative of a particular life zone, and can serve as an early warning sign of environmental stress over a wide range and intensity of stress.[3] Indicator species should also help diagnose the effects on that ecosystem, instead of merely identifying a change. Indicator species are a cornerstone of ecotoxicology, a discipline of toxicology that aims to understand the effect of toxicants on ecology in order to mitigate the effects of anthropogenic pressures on the environment.[4]

Omics technologies and systems biology (Figure 1.1) provides a global view of cellular processes on an individual to population-wide level, providing the opportunity to revolutionize how scientist use indicator species in biomonitoring. Stress exposure (e.g., pesticide, climate change) induces changes in gene expression and protein production, which are reflected in the small molecule metabolites that comprise an organism's metabolome.[1] Changes in metabolite levels can be used to indicate deviations from

homeostasis in an individual or population. This field of study, known as metabolomics, provides a promising avenue to examine the effects of contaminants on lower trophic level species and provide insights into environmental problems prior to food chain-wide effects.

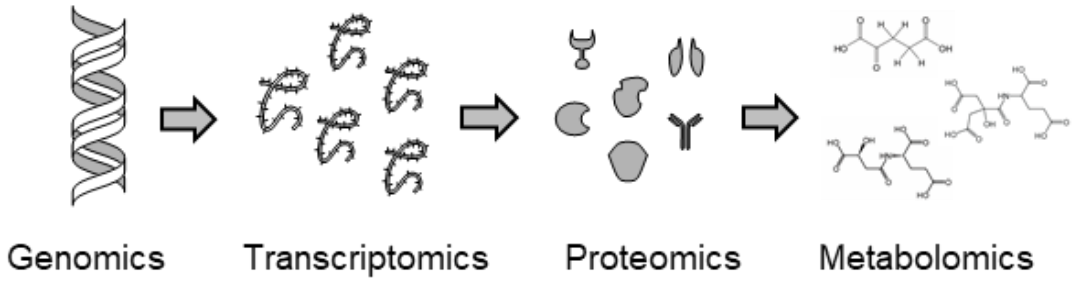


Figure 1.1. Depiction of the hierarchy of systems biology.

1.1.1. Metabolomics

Metabolomics is used to track global changes in metabolites in biofluids, cells, tissues, and organs from stressors, such as disease, toxicity, or environmental conditions.[1-2] The goal of metabolomics experiments is to quantify perturbations in the metabolome to provide signatures of an applied stress. Note the difference between xenobiotic metabolites produced through detoxification mechanisms, and endogenous metabolites, which are products or byproducts of normal metabolic pathways and are the focus of metabolomics studies. Metabolomics encompasses the study of all primary and secondary metabolites in an organism such as amino acids, lipids, organic acids, terpenoids, and sugars. Metabolomics methods can be targeted, which are designed to probe defined compounds with similar physiochemical properties or metabolic pathways, or untargeted (i.e., globally)

which aims to detect all known or unknown metabolites in a sample.[5] For example, lipidomics is a branch of metabolomics that specifically investigates lipid metabolism.[6]

Metabolomic studies can be further divided into metabolic profiling and metabolic fingerprinting. Although these terminologies are often used interchangeably, metabolic fingerprinting refers to the pattern of a spectrum or chromatogram that distinguishes between healthy versus unhealthy, and thus, hundreds to thousands of metabolites are detected without knowledge of their identity.[2, 7-9] Metabolic fingerprinting experiments tend to be high-throughput and aim to determine if there is a difference in the metabolomes of two populations. Metabolic profiling aims to simultaneously detect and quantify metabolites in a sample set.[9] Metabolite identification is crucial to these experiments because they aim to understand the differences in metabolite levels. Since a major goal of metabolic profiling is to understand perturbations in biochemical pathways, this method is typically limited to identifiable metabolites. Metabolomics is the comprehensive study of all metabolites within a cell, tissue, biofluid, or organism, which is analytically unachievable since current technology cannot identify every single metabolite in a given sample.[2, 10]

1.1.2. Environmental Metabolomics

Metabolic profiles revealed through mechanistic investigations of toxicity and disease spawn novel approaches to diagnose fitness and investigate environmental health. As the outcome of genomics, transcriptomics, and proteomics, metabolomics reflects what has happened in an organism rather than what is happening now or may occur in the future.[2]

Environmental metabolomics is an expanding field that includes identifying toxicity signatures of environmental stressors to understanding the roles of microorganisms in biogeochemical processes.[1, 11-13] In terms of ecotoxicology, environmental metabolomics aims to identify and use metabolic biomarkers of xenobiotics to measure individual, population, and ecological health.[1] Functional changes induced by exposure to xenobiotics are observed at the molecular, cellular, organ, individual, and population levels and describing these deviations in adverse outcome pathway (AOP) models is increasingly emphasized.[14] Metabolomics may be a useful tool in defining toxicant mode of action (MOA) and aiding to the description of AOPs; however, many challenges remain in implementing metabolomics in systems toxicology, including validating and predicting biomarkers, relating metabolic perturbations to critical outcomes, and mapping species-specific metabolic pathways.

Perturbations in bioenergetic processes are fundamental metabolic responses to stressors, which strive to provide extra energy to maintain homeostasis. These strategies are commonly used and documented in the study of tumors and carcinogenesis (i.e., the Warburg effect) where increased lactate levels are a signature of upregulation of energy production and the result of anaerobic glycolysis, pentose phosphate pathway, and glutaminolysis.[15-16] In glutaminolysis, for instance, glutamine is converted into glutamate and then converted into α-ketoglutarate and pushed into the TCA cycle, either by reaction with oxaloacetate to produce aspartate as a byproduct or pyruvate to produce alanine as a byproduct. Lactate reduction, alternatively, may indicate gluconeogenesis or its use in ATP production, as seen in Chinook salmon pre-smelt exposed to dispersed oil.[17] Also observed,

histidine downregulation can signify its catabolism into α-ketoglutarate to increase production of ATP. Furthermore, changes in ratios or concentrations of carnitine biosynthesis metabolites may also indicate perturbation of β-oxidation.[18] Due to the interwoven nature of TCA cycle metabolites throughout the biochemical map, there are many combinations that indicate different strategies to increase energy production, so it is critical to consider all metabolite changes when trying to determine mode of action of stressors.

1.1.3. Metabolic Flux Analysis

Metabolic flux or fluxomics is a facet of metabolomics that examines the flow of metabolites in biochemical pathways or changes in the rate of metabolism.[19-20] Metabolic profiling is useful for identifying changes in metabolite levels, but it does not give insight into the rate of metabolism. Metabolic flux cannot be directly measured, instead it requires the introduction of stable (e.g., ^{2}H, ^{13}C, ^{15}N) or radio isotopes into cells or tissues and observation of their metabolism over time. Metabolic flux analyses are stoichiometric in nature, where the goal is to observe transient changes in metabolites within a metabolic pathway.[21] All the common pathways and their corresponding enzymatic activity has been elucidated, so these experiments tend to be more targeted to probe the regulation of metabolites.[19] Isotopic enrichment is analyzed using mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR). MS measures mass, therefore MS cannot distinguish the position of the label and only detects the number of labeled nuclei present, which are known as isotopologues.[22] NMR spectroscopy can decipher the number and

position of isotopic labels, which are referred to as isotopomers.[22] Metabolic flux analyses are challenging but powerful tools that aid our understanding of the effects of stressors on metabolism.

1.2. Measuring Metabolic Perturbations

^{1}H NMR and MS are the most common analytical tools used to measure metabolites in biological samples.[9] NMR is an inherently quantitative, non-destructive, and highly reproducible analytical tool capable of rapidly detecting a wide range of metabolites with minimal sample preparation.[1-2, 23-25] Analysis by NMR is independent of solvent choice, polarity, and pK_a, thus offers a less biased analysis compared to other techniques.[1, 23] NMR offers a large dynamic range and is inherently quantitative, where peak intensity is directly proportional to concentration.[1, 23, 26] NMR experiments rely on signal averaging, where scans are acquired typically in multiples of four and summed together to increase the signal-to-noise ratio.[27] Many NMR-active nuclei are biologically relevant and include ^{1}H, ^{13}C, ^{15}N, and ^{31}P, allowing for detection and confirmation of metabolites using heteronuclear NMR techniques. Sensitivity of each nuclei is dependent on their gyromagnetic ratio (γ) and natural abundance, unless performing stable isotope enrichment experiments.[27] Sensitivity increases proportionally with magnetic field strength as γ^{5}; however, noise also increases by the square root of the observation frequency, meaning the signal-to-noise ratio scales as $\gamma^{5/2}$ with field strength.[27] Spectral dispersion also increases with magnetic field strength, resulting in sharper peaks and increased resolution. Cryogenically cooled probes and receivers have become more popular in routine NMR

spectroscopy because they reduce thermal noise (i.e., Johnson-Nyquist noise) resulting in a 4-8 times gain in sensitivity compared to conventional systems.[27] Though higher magnetic fields and longer acquisition times can enhance the sensitivity of NMR measurements,[23] using the most easily accessible instruments operating at 500 - 700 MHz and efficient acquisition times (10 – 60 min), ^{1}H NMR is limited to detection of micromolar or lower concentrations of many metabolites. Therefore an advantage of complementary analysis with MS is its generally greater sensitivity, detecting metabolites as low as picomolar concentrations.[5]

The coupling of MS detection with separation methods like GC or liquid chromatography (LC) offers further opportunities for untargeted profiling, such as characterizing the detoxification mechanisms of nano-Cu on cucumber plants, and targeted analysis, like profiling ascarylose-derived glycolipids in *Caenorhabditis elegans*.[28-31] GC-MS typically offers better separation and resolution than LC-MS, and effectively avoids ion suppression issues when using electron ionization; yet, for most metabolites GC-MS requires derivatization by silylation to increase metabolite volatility.[32] Additionally, large metabolite libraries exist for GC-MS and NMR, providing straightforward metabolite identification that is particularly useful in untargeted approaches. LC-MS is widely used in targeted analyses due to broad choices in suitable solvents, column chemistry, and polarity that allow researchers to develop strategies to screen for particular metabolites or pathways; however, LC-MS is still well-suited for untargeted approaches, although matrix effects and sample complexity can contribute to a poor LC separation.[33] Nevertheless, LC-MS is superior in detecting more metabolites compared to other techniques due to its high

sensitivity and chromatographic versatility (e.g., polarity, charge). One instrument cannot comprehensively measure every metabolite present in sample, therefore experiments that combine NMR and MS data can greatly expand metabolite coverage beyond what is possible using either technique alone.[28, 30, 34-39]

1.2.1. Sample Preparation

Sample preparation methods for metabolomics studies tend to be simple to extract as many metabolites as possible; however, it is a critical step that must be conducted carefully. Biological matrices vary (e.g., biofluids, cells, tissues, and whole-organism) and each sample requires optimization, but there are some general steps that can be taken to ensure high reproducibility. Generally, samples need to be quenched after removal from organisms to halt enzymatic activity.[1] Quenching with liquid nitrogen is suitable for cells, tissues, and whole-organism.[1, 40] Ice-cold methanol can also be used to quench metabolism in cells.[41-42] Some enzymes, like in the fungi species *Neurospora crassa*, are particularly robust and require heating to halt metabolism.[43-44]

For tissue samples, homogenization is often the next step which can be conducted before or after lyophilization.[1, 37, 45] Homogenization can be conducted by grinding tissue with a mortar and pestle with liquid nitrogen to keep it cold.[44] Additionally, tissue samples can be bead-beaten or homogenized with a spatula.[37, 46] The liquid or homogenized solid samples are subjected to extraction to solubilize metabolites, precipitate proteins and remove lipids, although this step is not always necessary. The most common extraction method is the Bligh and Dyer method, where 2:2:1.8 or 1:1:0.8 ice-cold

chloroform:methanol:water is used to separate polar and nonpolar metabolites.[47] Other common methods replace methanol with acetonitrile or isopropanol, or change solvent ratios.[5, 44-45] Perchloric acid extraction is another common extraction method used for polar metabolites, but should be used with care as some metabolites can be oxidized by this treatment.[45, 48] Single solvent extractions occasionally are performed with buffers, H_2O/D_2O, acetonitrile, or methanol.[1, 5, 46] Following extraction, the aqueous layer and, if performing lipidomics, nonpolar layer are dried under nitrogen gas or subjected to speedvac because dried extracts are more stable and stored at -80°C until analysis.

1.2.2. Nuclear Magnetic Resonance Spectroscopy

Nuclei with a nuclear spin quantum number (I) greater than zero possess nuclear spin.[27] When placed into a magnetic field (B_0), a magnetic moment (μ) is generated by the angular momentum (P) and charge. The gyromagnetic ratio (γ) is a constant unique for each nucleus and defines the strength of the magnetization.[27] The magnetic moment described by the equation $\mu = \gamma P$ and is defined as the torque the nucleus experiences as a result of the magnetic field. This torque causes the nuclei to precess when placed in the center of a static, external magnet. The rate of precession is known as the Larmor frequency, $\omega_0 = -\gamma B_0$ and the direction of the precession is defined by the sign of γ.

Protons (^1H), the most common nuclei detected in metabolomics experiments, have spin-½ and reside in two energy states. The lower energy state, α, is aligned with the applied magnetic field (B_0) and the higher energy state, β, is opposed to the applied field. Because the energy difference between the two states is small at thermal equilibrium, the

population difference between the two states as defined by the Boltzmann distribution is also small.[27, 49-50] As the energy difference between the spin states and therefore, their population difference, increases with applied magnetic field, NMR sensitivity also increases with magnetic field strength, roughly as $B_0^{3/2}$.[27] However, even at the highest available magnetic fields, this population difference is still small, making NMR less sensitive than many other spectroscopic techniques. Experiments such as dynamic nuclear polarization (DNP) increase the nuclear population difference by transfer of electron spin polarization and hold much promise for increasing both the sensitivity and the selectivity of metabolic profiling and flux experiments.[51-52]

The detected signal in pulsed NMR experiments is produced by application of radiofrequency (rf) pulse creating a transient secondary magnetic field B_1 orthogonal to B_0. The rf pulse perturbs the populations of the two spin states and creates coherence as the spins precess around both B_0 and B_1. This coherent precession induces an image current in the receiver coil of the probe that is recorded as a free induction decay (FID). Fourier transformation of the FID generates the NMR spectrum.

After the rf pulse, nuclei relax through two separate processes known as spin-lattice relaxation or longitudinal relaxation (T_1) and spin-spin or transverse relaxation (T_2). Spin-lattice relaxation leads to the re-establishment of equilibrium spin state populations. Obtaining accurate concentrations requires care to ensure that complete relaxation (i.e., return to equilibrium) occurs between scans or, if shorter relaxation delays are used, that correction factors based on the unique T_1 relaxation times of each nucleus are applied.[23, 52-53] Spin-spin relaxation causes the coherence generated by the rf pulse to be lost. For

small molecule metabolites in non-viscous solutions the T_2 relaxation times are typically in the range of 1-5 s.[27] Resonance line widths are inversely proportional to the apparent T_2 (T_2^*) which includes the natural T_2 and contributions from magnetic field inhomogeneity.

Metabolomic samples are complex mixtures of metabolites, leading to spectral complexity and there are important steps to take to yield high-quality NMR data. Shimming is a critical component of achieving a good NMR spectrum by correcting for magnetic field inhomogeneities to achieve a thin, sharp, and symmetrical lineshape.[27] Additionally, tuning and matching of the probe head is required to compensate for sample-dependent differences in conductivity and dielectric constant and match the resonance frequency of the rf coil to the Larmor frequency of the nuclei to be detected.[27, 50] Metabolomic samples, even those reconstituted in deuterated solvent, often contain some amount of water, so solvent suppression methods are used to reduce the intensity of the water signal and maximize the receiver gain. There are many types of water suppression methods, with the most commonly used in metabolomics being presaturation which applies a radiofrequency field centered around the resonance of water directly before acquisition, saturating the water's spins so that they are no longer observable.[27, 54]

Though the Larmor equation predicts that all nuclei of a given type would have the same spin, in practice even a simple molecule typically gives rise to an NMR spectrum with multiple resonances. The circulating electrons around the nucleus create a secondary induced magnetic field (B_{loc}) that opposes the applied magnetic field, partially shielding the nucleus from the effects of the applied field. The resonance frequency, ω, (called the chemical shift) of a given nucleus is dictated by its local environment ($\omega = -\gamma B_{loc}$). The

effect of shielding on the local magnetic field can be represented by introducing a constant, σ, which is unique for each individual chemical environment, $B_{loc} = B_0 (1 - \sigma)$. Electron donating groups increase the shielding of a nucleus (larger σ, lower chemical shift) while electronegative atoms or electron withdrawing groups like carboxylic acids or aldehydes lead to deshielding and a larger chemical shift.[49]

In addition to affecting chemical shift, neighboring NMR-active nuclei couple with each other through bonds increasing the multiplicity of the resonances and further complicating the spectrum. This type of coupling is known as scalar or *J*-coupling and it provides structural information about the number of NMR-active nuclei connected to a nucleus through chemical bonds. The multiplicity of resonances can be very simple with, no coupling (i.e., a singlet) and increase in complexity to doublets, triplets, quartets, doublet of doublets, etc. In cases where the difference in frequency between coupled nuclei is similar in magnitude to the coupling constant, the nuclei are said to be strongly coupled and both the multiplicity and the relative intensity of the resonances is more complex and must be analyzed through quantum mechanics equations that account for mixing of the spin states.[27]

For weakly coupled spin ½ nuclei, the multiplicity and intensity of their resonances is related by Pascal's triangle. In a doublet the intensity of the peaks is in a 1:1 ratio, while in a triplet the peak intensities are in a 1:2:1 ratio, and so forth.[27, 49] The multiplicity of a resonance is determined by the relationship 2NI + 1, where N is the spin (e.g., ^{1}H = ½, ^{2}H =1), and I is the number of coupled nuclei. For instance, α-ketoglutarate has four protons (Figure 1.2), where each set of protons has two neighbors and therefore multiplicity =

2(½)2 + 1 = 3. As shown in the red spectrum of Figure 1.2, α-ketoglutarate has as spin matrix consisting of two triplets. The two resonances have distinct chemical shifts due to their different chemical environments where the carbonyl group (C_2) causes deshielding of the protons bonded to C_3 shifting the triplet downfield. Chemical shift and *J*-coupling are the basis for assigning resonances and identifying metabolites in complex metabolomics samples.

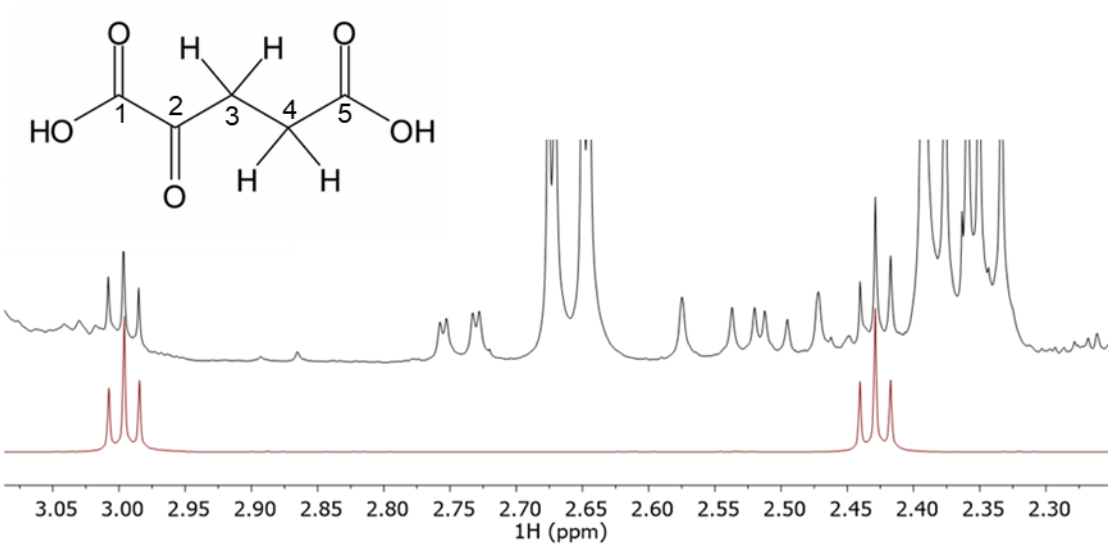


Figure 1.2. α-Ketoglutarate chemical structure and illustration of resonance assignment in the 1D NMR spectrum of a biological (black). The spectrum of an authentic standard of α-ketoglutarate (red) is plotted below the sample spectrum. The resonance chemical shifts and *J*-coupling patterns of the α-ketoglutarate standard align well with the sample resonances, confirming its assignment.

1.2.2.1. NMR Resonance Assignments

It is imperative to know what metabolites are present in a sample because changes in metabolite levels signal perturbations in biochemical pathways. Therefore, identification of metabolite resonances is a primary step in NMR-based metabolomics. An advantage of NMR is that it detects the resonances of all molecules simultaneously and in an unbiased manner. However, because metabolomics samples contain tens to hundreds of molecules above the NMR detection limit, spectra contain many resonances and resonance overlap is a common problem.

Spectral libraries like the Human Metabolome Database (HMDB), Chenomx Software, and the Madison Metabolomics Consortium Database are the first pass for the assignment of NMR resonances.[55-56] These databases are used to suggest the identity of metabolites and help make assignments. Because pH, temperature, and the sample matrix can affect the chemical shift and *J*-coupling of resonances, it is critical to purchase authentic standards of suspected metabolites, when available, and acquire spectra using the same buffer and experimental conditions used for the samples. For example, Figure 1.2 compares the NMR spectra of an α-ketoglutarate standard (red) and a biological sample (black). The chemical shifts and *J*-coupling of both triplets align well with the unknown resonances, suggesting that α-ketoglutarate is indeed the correct molecule.

Due to the complexity of biological samples and similarities in chemical shift and *J*-coupling of many metabolites, it is important to use two-dimensional (2D) NMR spectroscopy experiments to confirm assignments and elucidate the structures of unknown components. Experiments that provide homonuclear (e.g. ^{1}H-^{1}H) and heteronuclear

correlations (e.g. ^{1}H-^{13}C) are very helpful in confirming metabolite assignments. The simplest 2D pulse sequence, the [^{1}H-^{1}H] 2D *J*-resolved spectroscopy experiment, disperses chemical shift and coupling into two dimensions. This experiment is particularly useful for resolving resonances in overlapped regions of the spectra where it can be difficult to discern between a resonance's multiplicity.[27, 57-58] Also, because a 1D proton-decoupled spectrum can be generated by the F1 projection, overlapped multiplets can be resolved and integrated.

Correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) are two related homonuclear 2D NMR experiments with COSY providing correlations for directly coupled (i.e., neighboring) nuclei and the TOCSY provides correlations within spin systems (i.e., rings).[27, 59-62] The diagonal peaks in these experiments is equivalent to the 1D spectrum and the off-diagonal peaks, called cross-peaks, are correlations between neighboring spins. Following connections between the diagonal and cross-peaks facilitates resonance assignments especially for peaks that are obscured by overlap in the 1D spectrum.

The heteronuclear single quantum correlation spectroscopy or HSQC is the most common heteronuclear experiment used. This experiment displays correlations between ^{1}H and attached ^{13}C, providing an extra layer of confidence in the assignment since correlations of two different nuclei are assessed.[27, 63-66] In addition, the greater chemical shift dispersion of ^{13}C helps to overcome resonance overlap problems in complex mixtures. Lastly, the heteronuclear multiple bond correlation spectroscopy (HMBC) is a particularly useful experiment when trying to elucidate the structure of unknown compounds because

it shows correlations between ^{1}H and neighboring ^{13}C nuclei providing information about which groups are connected to each other. [27, 67]

1.2.2.2. Measuring Isotopic Enrichment

Measuring metabolic flux with NMR requires a different approach compared to metabolic profiling experiments. The most common isotopic labels used in flux experiments are ^{13}C and ^{15}N, both spin ½ nuclei with low levels of natural abundance. As described in section 1.2.2, coupling between two spin ½ nuclei leads to a doublet. For compounds with ^{13}C at natural abundance (1.1%) this leads to two resonances, called satellites on either side of the ^{1}H-^{12}C resonance at an intensity of 0.55% of the center peak. The one-bond coupling constant between the ^{1}H-^{13}C resonances is much larger than that observed between adjacent protons, typically 120 – 160 Hz.[27] Because of the low probability of having two adjacent 13C nuclei ($.011^{2}$), smaller long range couplings are not easily observed in ^{1}H NMR spectra of compounds with natural abundance levels of ^{13}C.

In^{13}C flux experiments, cells are provided with a precursor metabolite, for example glucose, incorporating the label at levels near 100%, either uniformly or as a site-specific label. As the ^{13}C-labeled glucose is metabolized, the ^{13}C nuclei are incorporated into downstream metabolites giving rise to more intense satellites on either side of the ^{1}H-^{12}C resonance than are observed in the natural abundance spectrum (Figure 1.3). Long range coupling to ^{13}C nuclei through 2 or 3 bonds gives the ^{1}H-^{13}C satellite peaks a more complex multiplet pattern than is observed for the ^{1}H-^{12}C doublet or the natural abundance satellites. This multiplicity provides isotopomer information that can be useful in the elucidation of

metabolic pathways especially when a site-specific label is used. It is possible to use 1D ^{13}C spectra to calculate positional enrichment of isotopomers which can be advantageous since ^{13}C is only 0.01% naturally abundant, and thus the observed signals are primarily due to enrichment.[68] However, ^{13}C is four-times less sensitive than ^{1}H and ^{13}C nuclei have much longer T_1 relaxation times than ^{1}H, thus this approach is not suitable for dilute samples.

The simplest experiment to determine the extent of label incorporation involves acquisition of a simple 1D ^{1}H NMR spectrum and comparison of the peak areas of the satellites against the unlabeled resonance to calculate percent incorporation, where $\%^{13}C = \Sigma(^{1}H\text{-}^{13}C \text{ Peak Area}) / (\Sigma(^{1}H\text{-}^{13}C \text{ Peak Area}) + {}^{1}H\text{-}^{12}C \text{ Peak Area})$.[68-69] Overlap of the satellite resonances with the ^{1}H resonances from other components of the sample makes this quite challenging. To remove the background due to overlapping resonances, a second 1D experiment can be acquired with ^{13}C-decoupled during acquisition by applying a radiofrequency field at the center of the ^{13}C spectrum, effectively removing the $^{1}H\text{-}^{13}C$ coupling. This spectrum can be compared with the coupled 1D spectrum, and the two spectra can be subtracted yielding a spectrum that contains only the ^{13}C satellite peaks which can be easily integrated (Figure 1.3).

Because metabolic flux analysis can be challenging using 1D spectra, 2D NMR is often employed as the second dimension reduces spectra complexity. The zero-quantum filtered (ZQF)-TOCSY is commonly used to study flux because it provides a method to measure the position of isotopomers in a molecule without the need for separation.[70-72] Alternatively, the heteronuclear ($^{1}H\text{-}^{13}C$) 2D *J*-resolved spectroscopy experiment can be

used to disperse heteronuclear *J*-coupling (i.e., ^{13}C satellites) into a second dimension (Figure 1.4), reducing spectral complexity and providing a means to quantify ^{13}C-labeled isotopomers.[69-70, 73-76] Slices in the F2 dimension can be taken, simplifying the integration of resonances to calculate the percent ^{13}C incorporation (Figure 1.4). There are many different approaches to measuring metabolic flux, but NMR is usually the preferred method when sensitivity is not an issue due to its ability to provide absolute quantitation and isotopomer position.

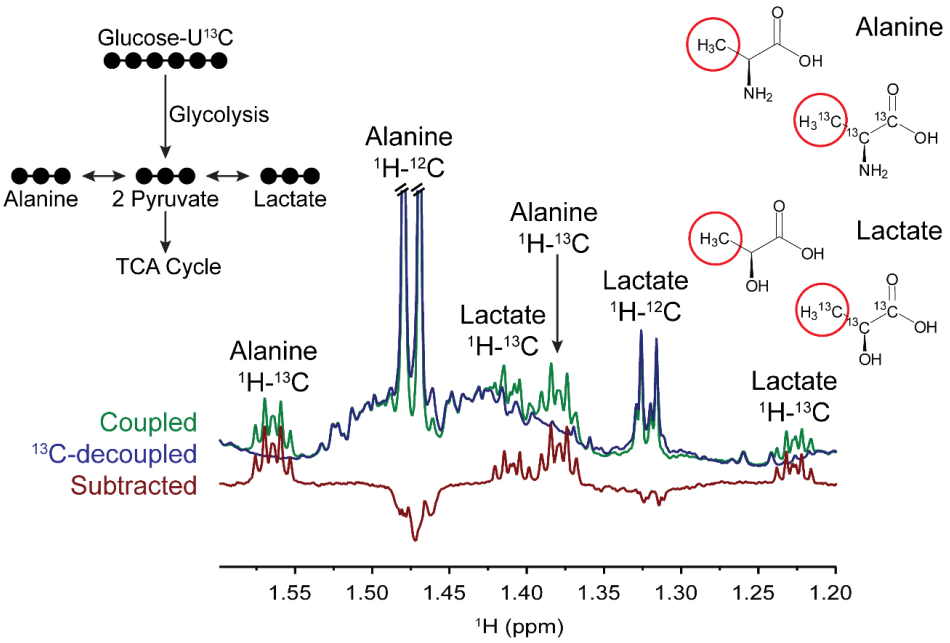


Figure 1.3. Representative 1D ^{13}C-coupled (green), ^{13}C-decoupled (blue), and subtracted (red) spectra of an extract of earthworm coelomocytes dosed with uniformly ^{13}C-labeled glucose. ^{13}C satellites are the resonances labeled with ^{1}H-^{13}C, while resonances labeled ^{1}H-^{12}C represent the resonances of the unlabeled fraction of alanine and lactate in the sample. The multiplicity of the ^{13}C satellites is indicative of uniformly labeled alanine and lactate.

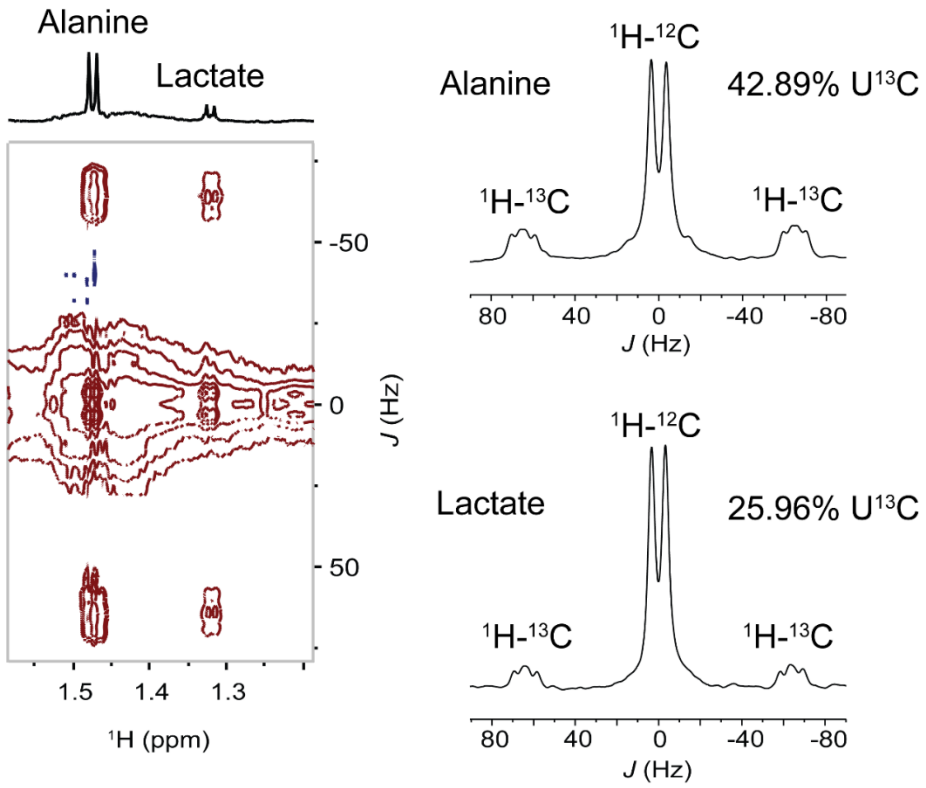


Figure 1.4. Representative heteronuclear 2D *J*-resolved spectrum (left) showing the dispersion of alanine and lactate ^{13}C satellite resonances into the second dimension, with 1D slices (right) showing the ^{1}H-^{12}C and ^{1}H-^{13}C resonances used to calculate percent enrichment for alanine and lactate.

1.2.3. Mass Spectrometry

Mass spectrometry (MS) offers complementary tools to expand and confirm biomarkers of exposure with greater sensitivity and resolution than NMR it requires small sample volumes.[1] Coupling a gas chromatography (GC) or liquid chromatography (LC) separation with MS detection provides the opportunity to target specific metabolites or pathways by adjusting the column, eluents or methods to favor physiochemical properties

of compounds, like hydrophobicity or volatility. Even in untargeted approaches, method biases such as column choice, ionization efficiency, and matrix effects must be considered. Although often preferred, separation is not necessary for metabolomics studies. Direct infusion-mass spectrometry (DI-MS) detects metabolites simultaneously as a sample is infused into the mass spectrometer.[77] DI-MS experiments require a high-resolution instrumentation to reduce overlap and assign peaks using accurate mass.

Mass spectrometers detect the mass to charge ratio, denotated as *m/z*. To be amenable to MS detection, molecules must be ionized using an ion source before being introduced into the mass spectrometer. The most commonly used source in LC-MS and DI-MS is electrospray ionization (ESI).[77-79] ESI uses high voltage to ionize compounds as they flow through a capillary, generating small, highly-charged droplets that are directed into the mass spectrometer.[78-79] Other methods such as atmospheric pressure chemical ionization (APCI) and atmospheric pressure photo ionization (APPI) can also be used.[79] Electron ionization (EI) is the most commonly used source in GC-MS. EI collides high energy electrons from an ion beam with gas-phase molecules, fragmenting the molecule into ions.[80] Other sources used in GC-MS include chemical ionization (CI), although its use is uncommon in metabolomics studies.[80]

After ionization, ions travel through ion optics to the mass analyzer and detector. Common mass analyzers used in metabolomics research include the single quadrupole (Q), triple quadrupole (QQQ), time of flight (TOF), quadrupole - time of flight (QTOF), Fourier-transform ion cyclotron (FT-ICR), and orbitrap. Many metabolites are structural isomers, and thus, have the same mass but their functional groups are arranged differently.

Because MS only detects *m/z*, structural isomers cannot be differentiated through the mass spectrum alone, which makes separation techniques important for the analysis of complex samples. Tandem mass spectrometry (MS/MS) experiments can be used to confirm identification of metabolites.[81] Tandem MS instruments first isolate a parent ion, fragment that ion in a collision cell, and detect the resulting fragments.[81] The resulting fragmentation patterns can be pieced together like a puzzle to elucidate the parent ion structure.

1.2.3.1. Gas Chromatography – Mass Spectrometry

GC-MS is a robust, economical tool for detecting a broad range of metabolites. GC columns separate compounds based on volatility and properties like hydrophobicity or aromatic ring content, yet many metabolites do not meet the criteria of being thermally stable and semi-volatile.[32, 82-83] Derivatization methods have been developed to lower boiling points and increase stability, targeting amine, carboxylate, and hydroxyl functional groups.[32, 35, 82-83] The most common derivatization method is a two-step approach where methoxyamine is first added to protect the carbonyl, followed by a silylation step that reacts with amine, carboxylate, and hydroxyl groups to increase volatility (Figure 1.5).[28, 32, 39, 82-84] MSTFA (*N*-methyl-*N*-(trimethylsilyl)trifluoriacetamide) and BSTFA (bis(trimethylsilyl)trifluoroacetamide) add trimethylsilyl- groups and are the most common silylation reagents.[32, 82-83] Alternatively, *tert*-butyldimethylsilylation with *N*-(*tert*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (MTBSTFA) is particularly useful for structural elucidation and metabolic flux analysis since it yields a single parent ion, unlike the trimethylsilyl derivatives.[32, 83, 85-87] Other derivatization reagents have been employed

in metabolomic studies, such as hexamethyldisilazane (HMDS) and trifluoroacetic acid (TFA), but are less commonly used.[35, 88]

A significant advantage of trimethylsilyl derivatization is that large libraries exists to aid identification of unknown metabolites, such as the U.S. National Institute of Standards (NIST) Mass Spectral Library and the Golm Metabolome Database.[89] In addition to containing fragmentation patterns for metabolites, these libraries provide retention indices that significantly aid in compound identification in GC-MS. The retention index (RI) is calculated using relative retention times that compensate for differences in column and method, as long as the same column chemistry and a linear temperature gradient are used.[83, 90-92] Because GC methods are robust, RIs are a convenient aid in compound identification and make GC-MS particularly attractive for metabolomics research.

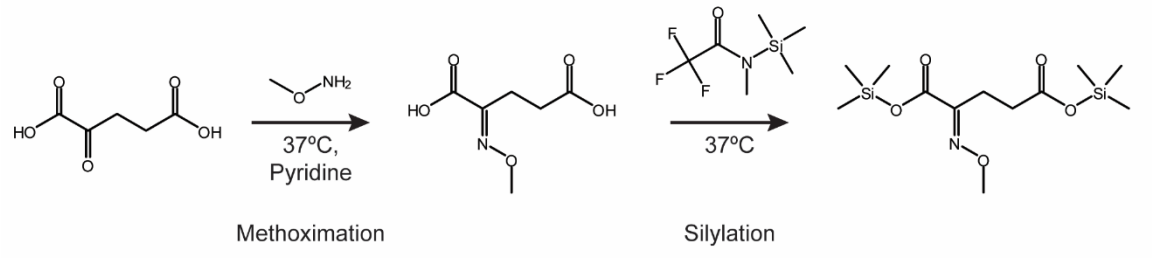


Figure 1.5. Derivatization scheme of α-ketoglutarate with methoxyamine and MSTFA for GC-MS analysis.

1.2.3.2. Liquid Chromatography – Mass Spectrometry

There are several chromatography strategies used in LC-MS metabolomic analyses, with the most common being reverse phase (RPLC) and hydrophilic interaction chromatography (HILIC).[5, 9] RPLC separates compounds based on polarity using a hydrophobic column (e.g. C18) and a gradient mobile phase, consisting typically of water and acetonitrile or methanol. Additives like buffers, acetic acid, and formic acid are often incorporated in the mobile phases to increase ionization efficiency and ion pairing reagents (e.g., trifluoroacetic acid, trimethylamine) may be used to improve the separation.[5, 33, 93] RPLC is capable of separating a broad spectrum of metabolites; however, obtaining good chromatographic resolution can be challenging for polar metabolites due to poor retention.[33] HILIC has gained in popularity because of the increased attention of polar metabolites, and is often used as complementary method to RPLC.[5, 33, 93-97] HILIC methods use polar columns with aprotic mobile phases, typically acetonitrile, with at least 5% water.[5, 94-95] A diverse collection of HILIC columns are available but are typically silica-based with acidic, basic, neutral, or zwitterionic functional groups.[5, 33, 96-99] The mechanism of HILIC separation is more complex than RPLC, and arises from a combination of chemical interactions, hydrophobic interactions, physical interactions, and Van der Waals forces.[94, 98] Identification of metabolites can be more challenging with LC-MS compared to NMR and GC-MS, but databases like the Metlin and XCMS libraries can be used with authentic standards and MS/MS to identify metabolites.[25, 100-103]

1.3. Chemometric Analyses

Statistical methods are needed to ascertain the relevant chemical information from the large datasets acquired in metabolomics. This approach is referred to as chemometrics, which heavily relies on multivariate analysis to reduce the complexity of the data.[104-105] Univariate analysis also plays an important role, although it can only relate a single independent variable to a single dependent variable.[104-105] Biological processes are multivariate in nature, thus multivariate approaches are often more useful because many variables are considered simultaneously.[104-105] It is typically good practice to perform both types of analyses because they yield different information and are complementary for identifying biomarkers.[105]

1.3.1. Univariate Approaches

Univariate statistical analysis is conducted with *t* tests for two group comparisons or analysis of variance (ANOVA) for comparisons of three or more groups. These methods test the equality of means among groups to within-group variance.[105-107] Prior to analysis, preprocessing is conducted to meet ANOVA assumptions such as normality and homoscedasticity. Sharpiro-Wilks tests and histograms are used to test normality and Levene's test is used to assess homoscedasticity.[108-109] Logarithm, power, box-cox, or other transformations can be used to correct the data to meet the assumptions. If transformations do not correct normality, nonparametric tests like the Kruskal-Wallis ANOVA, or if the data is normally distributed but heteroscedastic, the Welch ANOVA can be used.[109-110] Following the ANOVA, *post hoc* tests are used to assess significance such as Tukey's HSD

for the ANOVA or Games-Howell for the Welch ANOVA.[107, 111] Following significance assessment, fold change or effect sizes are calculated to evaluate the magnitude of changes in metabolite levels.[109, 112-113]

1.3.2. Multivariate Approaches

A suite of multivariate tools is used in metabolomics to understand correlations or covariances between variables (i.e., metabolites). These methods evaluate all variables simultaneously, unlike univariate which compares a single variable's mean and variance.[105] Prior to data reduction, raw data is preprocessed using centering, scaling, and transformations to reduce effects of noise, differences in metabolites concentration, biovariability, and heteroscedasticity.[114] Centering converts all metabolite fluctuations around zero instead of the mean prior to scaling, while scaling is used to adjust fold differences in metabolite concentrations by dividing each variable by a scaling factor. There are many types of scaling but the most commonly used in metabolomics include: autoscaling, pareto scaling, range scaling, vast scaling, and level scaling, with auto- and pareto scaling being the most common. Autoscaling or unit variance scaling divides variables by their standard deviation, while pareto scaling uses the square root the standard deviation as the scaling factor.[114] Transformation with power or logarithmic functions are used to correct heteroscedasticity. Lastly, multiblocking (MB) techniques allow for integration of datasets acquired with multiple instruments, for example NMR and MS, in chemometric analysis, encouraging the use of orthogonal analytical techniques to assess fitness.[24, 115]

Principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and orthogonal partial least squares-discriminant analysis (OPLS-DA) are the most common multivariate approaches used in metabolomics to visualize groupings and identify the metabolite drivers of group separation. PCA is an unsupervised method, meaning group identification is not known by the model, and is the most common approach to identify the metabolites that are the larger drivers of sample segregation.[104-105, 116-117] PLS-DA and OPLS-DA are supervised multivariate analysis, meaning group identification is known by the model, and are used to identify metabolites most important to the separation of groups. OPLS-DA introduces an orthogonal component that filters variance unrelated to group separation in an orthogonal component; however, it has no predictive advantage over PLS-DA.[116-117]

As observed in Figure 1.6, multivariate analysis yields score plots (left) and loadings plots (right). Score plots provide visualization of class separation, while loadings plots provide visualization of the drivers of class separation. For the metabolomics data presented in Figure 1.6, the score plots calculated using each multivariate method show good separation between groups, and the loadings plots indicate similar metabolite drivers for the separation. In all models, acetylcarnitine, betaine, nicotinamide, and glutamine contribute to separation of the red triangles (treatment) from the green circles (control). Arginine, hypoxanthine, leucine, methionine, phenylalanine, proline, trimethyllysine, and valine contributed to separation of the treatment samples (red circles) from the controls (green triangles) (Figure 1.6). Though in this example the loadings plots yield the similar results, the position of the variables is different in each model which is reflective of the

method to perform the analysis (i.e., PCA, PLS-DA, OPLS-DA). Multivariate analysis is important in the interpretation of metabolomics data because it allows identification of metabolite-group associations.

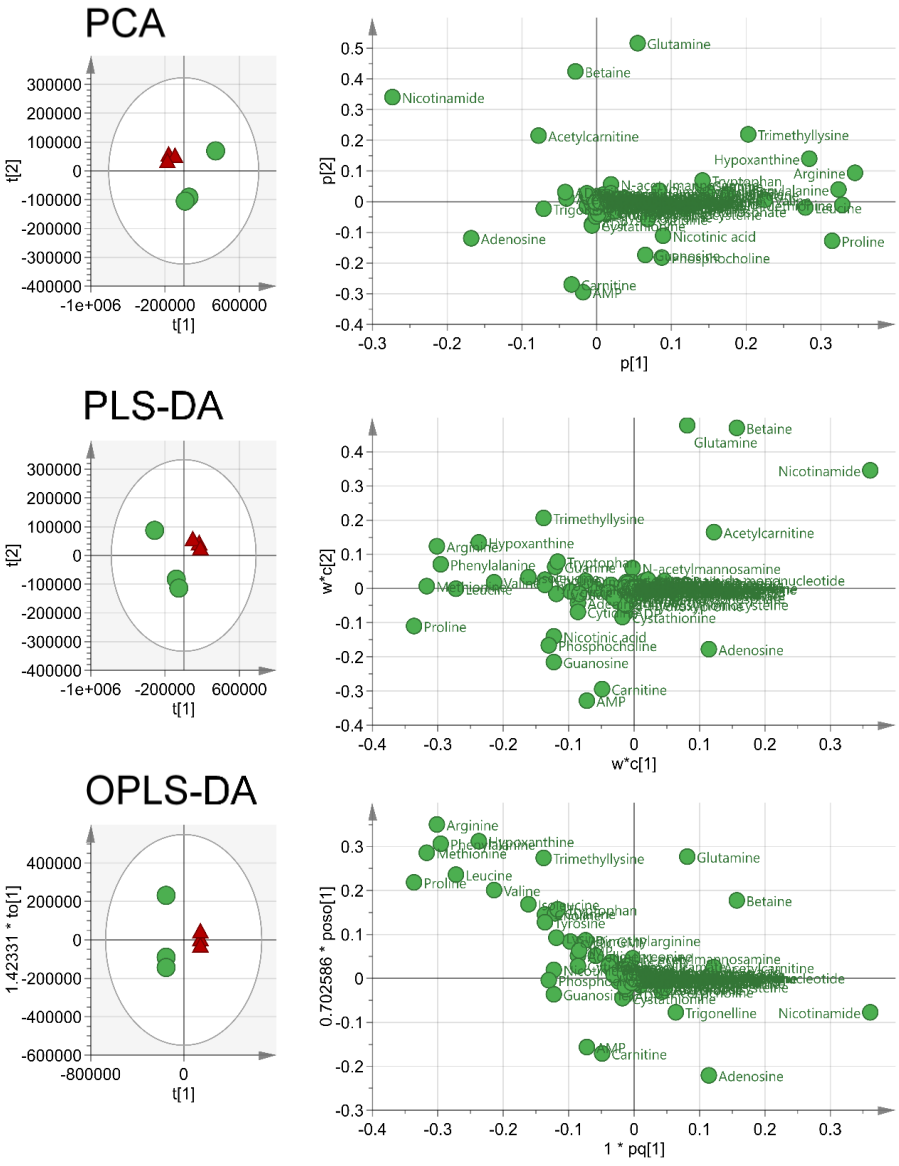


Figure 1.6. Representative score plots (left) and loadings plots (right) comparing control (green circle) and treatment (red triangle) samples groups by PCA, PLS-DA, and OPLS-DA. LC-MS data of earthworm extracts was used to generate these plots.

1.4. Earthworm Metabolomics

As detritivores and ecological engineers, earthworms are at the forefront of soil health and an obvious choice for environmental monitoring. Studies to assess the impacts of heavy metals, persistent contaminants, and pesticides provide evidence for the potential of using earthworm metabolomics as an indicator of ecological stress.[1] Proteomic and transcriptomic strategies have also been assessed as potential methods to monitor earthworm health.[118] Metabolomic and proteomic approaches were used in a prior study to examine the systemic impact of polybrominated diphenyl ether 47 (PBDE 47) on *Eisenia fetida* (Figure 1.7).[119] Evidence of perturbation of energy metabolism was observed in both data sets as a result of exposure to PBDE 47, while metabolomics also indicated disruption in osmoregulation and the proteomics results identified changes consistent with oxidative stress, apoptosis, and impeded protein synthesis. In a study of copper toxicity on *Lumbricus rubellus*, disruption of carbohydrate metabolism was observed in the transcriptomics and metabolomics profiling data.[120]

Though the metabolomic impacts of toxicants on earthworms have been probed with both ^{1}H NMR and GC-MS, more commonly analyses have been conducted independently. For example, GC-MS and ^{1}H NMR revealed that pyrene exposure increased the levels of several amino acids and decreased lactate and saturated fatty acids in the earthworm species *Lumbricus rubellus*.[34] In another study, deviations from the homeostatic alanine to glycine ratio of 1.5 were indicative of dichlorodiphenyltrichloroethane (DDT) and endosulfan exposure, as determined in the ^{1}H NMR and GC-MS whole-earthworm (*E. fetida*) profiles.[35] Furthermore, earthworms (*L. rubellus*) exposed simultaneously to

imidacloprid and thiacloprid, neonicotinoids with a similar pesticide MOA, revealed metabolic responses that suggested the compounds worked independently to perturb the metabolome, instead of compounding their effects as hypothesized.[36] The same study also examined how simultaneous exposure to the toxicants chlorpyrifos and nickel with dissimilar MOAs, impacted the earthworm metabolome, finding the metabolic effects to be independent from one another, as hypothesized.[36]

Figure 1.7. Picture of the earthworm, *Eisenia fetida,* (top) used in this research and the Magic Worm Ranch used to culture the worms in the laboratory.

Metabolic studies have largely focused on whole-worm extracts, but have also been extended to coelomic fluid (CF), a yellow liquid (Figure 1.8) that fills the worm's coelom and connects to the environment through a dorsal pore in each of its segments.[1, 121] Despite

these pores, the earthworm's coelom avoids sepsis, leading many researchers to hunt for unique defense proteins and peptides within the CF and to better understand the role of coelomocytes (CC) in combating pathogens and parasites.[122] CF holds cytotoxic, hemolytic, proteolytic, and antimicrobial defense mechanisms and plays critical roles in immunity, movement, excretion, nutrient storage, and metabolism.

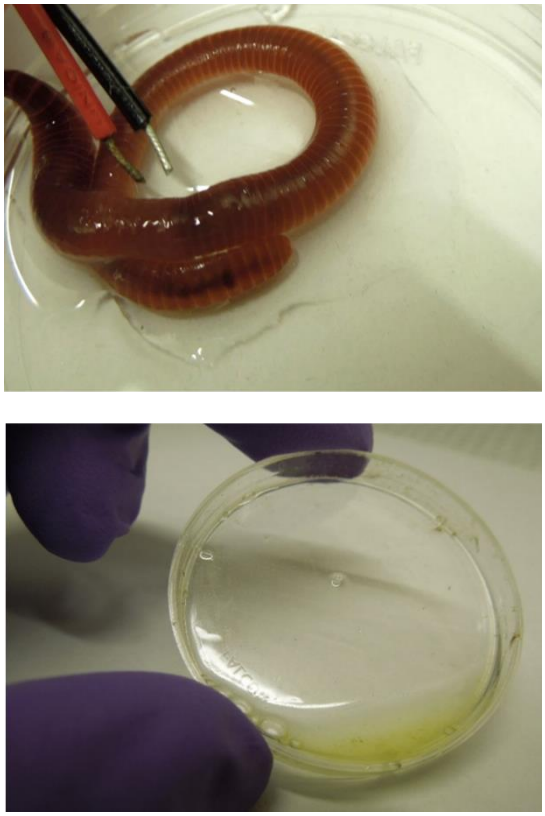


Figure 1.8. Demonstration of the nonlethal and noninvasive earthworm coelomic fluid and coelomocyte extrusion by applying an electrical impulse across the worm (top)[4], yielding the extrusion of the yellow biofluid (bottom).[123]

In assessing the effects of exposure to a contaminant, CF offers several advantages over whole-worm extracts including nonlethal sampling (Figure 1.8), simple sample preparation, and greater temporal resolution.[123] Although typically discarded for metabolomic studies, CC are free-moving cells within the CF that serve immunological and hepatocytic functions. To our knowledge, the CC metabolome has not been profiled or used in metabolomic studies, though it has been used to assess toxicity. For example, Burch, et al. [124] described an *in vitro* CC immunotoxicity assay, while others have used the comet and other genotoxic assays to evaluate genotoxicity.[125] Therefore, analysis of earthworm CC metabolites may provide an additional source of information to augment these toxicity assessments.

These studies highlight the potential of systems biology approaches, including metabolomics and metabolic profiling, to interpret the biochemical changes resulting from environmental insults, and motivated our efforts to seek a more comprehensive analysis of the *E. fetida* metabolome. The experiments presented in the dissertation aim to elucidate metabolite structures, understand earthworm metabolism, and probe the utility of coelomic fluid and coelomocytes in ecotoxicological studies.

1.5. References

1. Lankadurai, B. P.; Nagato, E. G.; Simpson, M. J., Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environ. Rev.* **2013,** *21* (3), 180-205.

2. Larive, C. K.; Barding, G. A.; Dinges, M. M., Nmr spectroscopy for metabolomics and metabolic profiling. *Anal. Chem.* **2015,** *87* (1), 133-146.

3. Carignan, V.; Villard, M.-A., Selecting indicator species to monitor ecological integrity: A review. *Environ Monit Assess* **2002,** *78* (1), 45-61.

4. Zhang, X.; Xia, P.; Wang, P.; Yang, J.; Baird, D. J., Omics advances in ecotoxicology. *Environ Sci Technol* **2018,** *52* (7), 3842-3851.

5. Cajka, T.; Fiehn, O., Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* **2016,** *88* (1), 524-545.

6. Yang, K.; Han, X., Lipidomics: Techniques, applications, and outcomes related to biomedical sciences. *Trends Biochem. Sci.* **2016,** *41* (11), 954-969.

7. Kopka, J.; Fernie, A.; Weckwerth, W.; Gibon, Y.; Stitt, M., Metabolite profiling in plant biology: Platforms and destinations. *Genome Biol.* **2004,** *5* (6), 109-109.

8. Dettmer, K.; Hammock, B. D., Metabolomics--a new exciting field within the "omics" sciences. *Environ. Health Perspect.* **2004,** *112* (7), A396-A397.

9. Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S., Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatogr. A* **2015,** *1382* (0), 136-164.

10. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for nmr spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protocols* **2007,** *2* (11), 2692-2703.

11. Bundy, J.; Davey, M.; Viant, M., Environmental metabolomics: A critical review and future perspectives. *Metabolomics* **2009,** *5* (1), 3-21.

12. Simpson, M. J.; McKelvie, J. R., Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. *Anal. Bioanal. Chem.* **2009,** *394* (1), 137-149.

13. Jones, O. A. H.; Lear, G.; Welji, A. M.; Collins, G.; Quince, C., Community metabolomics in environmental microbiology. In *Microbial metabolomics: Applications*

in clinical, environmental, and industrial microbiology, Beale, D. J.; Kouremenos, K. A.; Palombo, E. A., Eds. Springer International Publishing: Cham, 2016; pp 199-224.

14. Sturla, S. J.; Boobis, A. R.; FitzGerald, R. E.; Hoeng, J.; Kavlock, R. J.; Schirmer, K.; Whelan, M.; Wilks, M. F.; Peitsch, M. C., Systems toxicology: From basic research to risk assessment. *Chem. Res. Toxicol.* **2014,** *27* (3), 314-329.

15. Vaitheesvaran, B.; Xu, J.; Yee, J.; Lu, Q. Y.; Go, V. L.; Xiao, G. G.; Lee, W. N., The warburg effect: A balance of flux analysis. *Metabolomics* **2015,** *11* (4), 787-796.

16. Owen, O. E.; Kalhan, S. C.; Hanson, R. W., The key role of anaplerosis and cataplerosis for citric acid cycle function. *J. Bio. Chem.* **2002,** *277* (34), 30409-30412.

17. Van Scoy, A. R.; Yu Lin, C.; Anderson, B. S.; Philips, B. M.; Martin, M. J.; McCall, J.; Todd, C. R.; Crane, D.; Sowby, M. L.; Viant, M. R.; Tjeerdema, R. S., Metabolic responses produced by crude versus dispersed oil in chinook salmon pre-smolts via nmr-based metabolomics. *Ecotox. Environ. Safe.* **2010,** *73* (5), 710-717.

18. Wünschiers, R.; Jahn, M.; Jahn, D.; Schomburg, I.; Peifer, S.; Heinzle, E.; Burtscher, H.; Garbe, J.; Steen, A.; Schobert, M.; Oesterhelt, D.; Wachtveitl, J.; Chang, A., Metabolism. In *Biochemical pathways*, John Wiley & Sons, Inc.: 2012; pp 37-209.

19. Sauer, U., Metabolic networks in motion: 13c-based flux analysis. *Mol. Syst. Biol*, **2006,** *2* (1), 1-10.

20. Jang, C.; Chen, L.; Rabinowitz, J. D., Metabolomics and isotope tracing. *Cell* **2018,** *173* (4), 822-837.

21. Varma, A.; Palsson, B. O., Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology* **1994,** *12*, 994.

22. Buescher, J. M.; Antoniewicz, M. R.; Boros, L. G.; Burgess, S. C.; Brunengraber, H.; Clish, C. B.; DeBerardinis, R. J.; Feron, O.; Frezza, C.; Ghesquiere, B.; Gottlieb, E.; Hiller, K.; Jones, R. G.; Kamphorst, J. J.; Kibbey, R. G.; Kimmelman, A. C.; Locasale, J. W.; Lunt, S. Y.; Maddocks, O. D. K.; Malloy, C.; Metallo, C. M.; Meuillet, E. J.; Munger, J.; Nöh, K.; Rabinowitz, J. D.; Ralser, M.; Sauer, U.; Stephanopoulos, G.; St-Pierre, J.; Tennant, D. A.; Wittmann, C.; Vander Heiden, M. G.; Vazquez, A.; Vousden, K.; Young, J. D.; Zamboni, N.; Fendt, S.-M., A roadmap for interpreting 13c metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.* **2015,** *34*, 189-201.

23. Barding, G. A.; Salditos, R.; Larive, C. K., Quantitative nmr for bioanalysis and metabolomics. *Anal. Bioanal. Chem.* **2012,** *404* (4), 1165-1179.

24. Nagana Gowda, G. A.; Raftery, D., Recent advances in nmr-based metabolomics. *Anal. Chem.* **2017,** *89* (1), 490-510.

25. Wolfender, J.-L.; Nuzillard, J.-M.; van der Hooft, J. J. J.; Renault, J.-H.; Bertrand, S., Accelerating metabolite identification in natural product research: Toward an ideal combination of lc-hrms/ms and nmr profiling, in silico databases and chemometrics. *Anal. Chem.* **2018**.

26. Bearden, D. W., Environmental metabolomics. In *Emagres*, John Wiley & Sons, Ltd: 2007.

27. Claridge, T. D. W., *High-resolution nmr techniques in organic chemistry*. 1 ed.; Elsevier: Oxford, UK, 1999; Vol. 19.

28. Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K., Comparison of gc-ms and nmr for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12* (2), 898-909.

29. Mathon, C.; Barding Jr, G. A.; Larive, C. K., Separation of ten phosphorylated mono-and disaccharides using hilic and ion-pairing interactions. *Anal. Chim. Acta* **2017,** *972*, 102-110.

30. Zhao, L.; Huang, Y.; Hu, J.; Zhou, H.; Adeleye, A. S.; Keller, A. A., 1h nmr and gc-ms based metabolomics reveal defense and detoxification mechanism of cucumber plant under nano-cu stress. *Environ. Sci. Technol.* **2016,** *50* (4), 2000-2010.

31. von Reuss, S. H.; Dolke, F.; Dong, C., Ascaroside profiling of caenorhabditis elegans using gas chromatography–electron ionization mass spectrometry. *Anal. Chem.* **2017,** *89* (19), 10570-10577.

32. Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T., Quantitative metabolomics based on gas chromatography mass spectrometry: Status and perspectives. *Metabolomics* **2011,** *7* (3), 307-328.

33. Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D., Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* **2014,** *87*, 12-25.

34. Jones, O. A. H.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L., A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm lumbricus rubellus. *Chemosphere* **2008,** *71* (3), 601-609.

35. McKelvie, J. R.; Yuk, J.; Xu, Y. P.; Simpson, A. J.; Simpson, M. J., H-1 nmr and gc/ms metabolomics of earthworm responses to sub-lethal ddt and endosulfan exposure. *Metabolomics* **2009,** *5* (1), 84-94.

36. Baylay, A. J.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L.; Swain, S.; Sturzenbaum, S.; Jones, O. A. H., A metabolomics based test of independent action and

concentration addition using the earthworm lumbricus rubellus. *Ecotoxicology* **2012,** *21* (5), 1436-1447.

37. Griffith, C. M.; Williams, P. B.; Tinoco, L. W.; Dinges, M. M.; Wang, Y.; Larive, C. K., 1h nmr metabolic profiling of earthworm (eisenia fetida) coelomic fluid, coelomocytes, and tissue: Identification of a new metabolite – malylglutamate. *J. Proteome Res.* **2017,** *16* (9), 3407-3417.

38. Dinges, M. M.; Lytle, C.; Larive, C. K., 1h nmr-based identification of intestinally absorbed metabolites by ussing chamber analysis of the rat cecum. *Anal. Chem.* **2018**.

39. Griffith, C. M.; Morgan, M. A.; Dinges, M. M.; Mathon, C.; Larive, C. K., Metabolic profiling of chloroacetanilide herbicides in earthworm coelomic fluid using (1)h nmr and gc-ms. *J. Proteome Res.* **2018,** *17* (8), 2611-2622.

40. Lorenz, M. A.; Burant, C. F.; Kennedy, R. T., Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. *Anal. Chem.* **2011,** *83* (9), 3406-3414.

41. Canelas, A. B.; Ras, C.; ten Pierick, A.; van Dam, J. C.; Heijnen, J. J.; van Gulik, W. M., Leakage-free rapid quenching technique for yeast metabolomics. *Metabolomics* **2008,** *4* (3), 226-239.

42. Kostidis, S.; Addie, R. D.; Morreau, H.; Mayboroda, O. A.; Giera, M., Quantitative nmr analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Anal. Chim. Acta* **2017,** *980*, 1-24.

43. Ouyang, X.; Leonards, P. E. G.; Tousova, Z.; Slobodnik, J.; de Boer, J.; Lamoree, M. H., Rapid screening of acetylcholinesterase inhibitors by effect-directed analysis using lc × lc fractionation, a high throughput in vitro assay, and parallel identification by time of flight mass spectrometry. *Anal. Chem.* **2016,** *88* (4), 2353-2360.

44. Liebeke, M.; Bundy, J. G., Tissue disruption and extraction methods for metabolic profiling of an invertebrate sentinel species. *Metabolomics* **2012,** *8* (5), 819-830.

45. Kaiser, K. A.; Barding, G. A.; Larive, C. K., A comparison of metabolite extraction strategies for h-1-nmr-based metabolic profiling using mature leaf tissue from the model plant arabidopsis thaliana. *Magn. Reson. Chem.* **2009,** *47*, S147-S156.

46. Brown, S. A. E.; Simpson, A. J.; Simpson, M. J., Evaluation of sample preparation methods for nuclear magnetic resonance metabolic profiling studies with eisenia fetida. *Environ. Toxicol. Chem.* **2008,** *27* (4), 828-836.

47. Bligh, E. G.; Dyer, W. J., A rapid method of total lipid extraction and purification. *Can. J. Biochem. Phys.* **1959,** *37* (8), 911-917.

48. Belle, J. E. L.; Harris, N. G.; Williams, S. R.; Bhakoo, K. K., A comparison of cell and tissue extraction techniques using high-resolution 1h-nmr spectroscopy. *NMR Biomed.* **2002,** *15* (1), 37-44.

49. Akitt, J. W., *Nmr and chemistry: An introduction to the fourier transform-multinuclear era*. 2 ed.; Chapman and Hall: London, UK, 1983.

50. Hornak, J., *The basics of nmr*. 2017.

51. Barnes, A. B.; De Paëpe, G.; van der Wel, P. C. A.; Hu, K.-N.; Joo, C.-G.; Bajaj, V. S.; Mak-Jurkauskas, M. L.; Sirigiri, J. R.; Herzfeld, J.; Temkin, R. J.; Griffin, R. G., High-field dynamic nuclear polarization for solid and solution biological nmr. *Appl. Magn. Reson.* **2008,** *34* (3), 237-263.

52. Giraudeau, P., Challenges and perspectives in quantitative nmr. *Magn. Reson. Chem.* **2017,** *55* (1), 61-69.

53. Bharti, S. K.; Roy, R., Quantitative 1h nmr spectroscopy. *Trends in Anal. Chem.* **2012,** *35*, 5-26.

54. Giraudeau, P.; Silvestre, V.; Akoka, S., Optimizing water suppression for quantitative nmr-based metabolomics: A tutorial review. *Metabolomics* **2015**, 1-15.

55. Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the madison metabolomics consortium database. *Nat Biotech* **2008,** *26* (2), 162-164.

56. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., Hmdb 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **2013,** *41* (D1), D801-D807.

57. Aue, W. P.; Karhan, J.; Ernst, R. R., Homonuclear broad band decoupling and two-dimensional j-resolved nmr spectroscopy. *J. Chem. Phys.* **1976,** *64* (10), 4226-4227.

58. Huang, Y.; Zhang, Z.; Chen, H.; Feng, J.; Cai, S.; Chen, Z., A high-resolution 2d j-resolved nmr detection technique for metabolite analyses of biological samples. *Sci. Rep.* **2015,** *5*, 8390.

59. Piantini, U.; Sorensen, O. W.; Ernst, R. R., Multiple quantum filters for elucidating nmr coupling networks. *J. Am. Chem. Soc.* **1982,** *104* (24), 6800-6801.

60. Bax, A.; Davis, D. G., Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985,** *65* (2), 355-360.

61. Davis, A. L.; Laue, E. D.; Keeler, J.; Moskau, D.; Lohman, J., Absorption-mode two-dimensional nmr spectra recorded using pulsed field gradients. *J. Magn. Reson.* **1991,** *94* (3), 637-644.

62. Liu, M.; Mao, X.-a.; Ye, C.; Huang, H.; Nicholson, J. K.; Lindon, J. C., Improved watergate pulse sequences for solvent suppression in nmr spectroscopy. *J. Magn. Reson.* **1998,** *132* (1), 125-129.

63. Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M., Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation nmr spectroscopy. *J. Magn. Reson.* **1991,** *93* (1), 151-170.

64. Kay, L.; Keifer, P.; Saarinen, T., Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992,** *114* (26), 10663-10665.

65. Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W., Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **1993,** *31* (3), 287-292.

66. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C., A general enhancement scheme in heteronuclear multidimensional nmr employing pulsed field gradients. *J. Biomol. NMR* **1994,** *4* (2), 301-306.

67. Bax, A.; Subramanian, S., Sensitivity-enhanced two-dimensional heteronuclear shift correlation nmr spectroscopy. *J. Magn. Reson.* **1986,** *67* (3), 565-569.

68. Masakapalli, S. K.; Ratcliffe, R. G.; Williams, T. C. R., Quantification of 13c enrichments and isotopomer abundances for metabolic flux analysis using 1d nmr spectroscopy. In *Plant metabolic flux analysis: Methods and protocols*, Dieuaide-Noubhani, M.; Alonso, A. P., Eds. Humana Press: Totowa, NJ, 2014; pp 73-86.

69. Aue, W. P., Karhann J., Ernst, R. R., Homonuclear broad band decoupling and two-dimensional j-resolved nmr spectroscopy. *J. Chem. Phys.* **1976,** *64* (10), 4226-4227.

70. Thrippleton, M. J.; Keeler, J., Elimination of zero-quantum interference in two-dimensional nmr spectra. *Angew. Chem.* **2003,** *42* (33), 3938-3941.

71. Massou, S.; Nicolas, C.; Letisse, F.; Portais, J.-C., Nmr-based fluxomics: Quantitative 2d nmr methods for isotopomers analysis. *Phytochemistry* **2007,** *68* (16), 2330-2340.

72. Massou, S.; Nicolas, C.; Letisse, F.; Portais, J.-C., Application of 2d-tocsy nmr to the measurement of specific 13c-enrichments in complex mixtures of 13c-labeled metabolites. *Metabolic Engineering* **2007,** *9* (3), 252-257.

73. Giraudeau, P.; Akoka, S., A new detection scheme for ultrafast 2d j-resolved spectroscopy. *J. Magn. Reson.* **2007,** *186* (2), 352-357.

74. Luy, B., Adiabatic z-filtered j-spectroscopy for absorptive homonuclear decoupled spectra. *J. Magn. Reson.* **2009,** *201* (1), 18-24.

75. Cahoreau, E.; Peyriga, L.; Hubert, J.; Bringaud, F.; Massou, S.; Portais, J.-C., Isotopic profiling of 13c-labeled biological samples by two-dimensional heteronuclear j-resolved nuclear magnetic resonance spectroscopy. *Anal. Biochem.* **2012,** *427* (2), 158-163.

76. Pathan, M.; Akoka, S.; Giraudeau, P., Ultrafast hetero-nuclear 2d j-resolved spectroscopy. *J. Magn. Reson.* **2012,** *214* (Supplement C), 335-339.

77. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007,** *26* (1), 51-78.

78. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **2013,** *85* (1), 2-9.

79. Awad, H.; Khamis, M. M.; El-Aneed, A., Mass spectrometry, review of the basics: Ionization. *Appl. Spectrosc. Rev.* **2015,** *50* (2), 158-175.

80. Harrison, A., *Chemical ionization mass spectrometry*. 2 ed.; Routledge: New York, 1992.

81. McLafferty, F., Tandem mass spectrometry. *Science* **1981,** *214* (4518), 280-287.

82. Xu, F.; Zou, L.; Ong, C. N.; Zou, L.; Ong, C. N.; Ong, C. N., Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based gc-ms metabolomics. *Trends in Anal. Chem.* **2010,** *29* (3), 269-280.

83. Fiehn, O., Metabolomics by gas chromatography-mass spectrometry: The combination of targeted and untargeted profiling. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2016,** *114*, 30.4.1-30.4.32.

84. Lee, D. Y.; Fiehn, O., High quality metabolomic data for chlamydomonas reinhardtii. *Plant Methods* **2008,** *4*, 13.

85. Baker, A.; Dodd, C. D.; Parsons, R., Identification of amino compounds synthesized and translocated in symbiotic parasponia. *Plant Cell Environ.* **1996,** *19* (11), 1249-1260.

86. Godber, I. M.; Parsons, R., Translocation of amino acids from stem nodules of sesbania rostrata demonstrated by gc-ms in planta 15n isotope dilution. *Plant Cell Environ.* **1998,** *21* (11), 1089-1099.

87. Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L., Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000,** *72* (15), 3573-3580.

88. Katona, Z. F.; Sass, P.; Molnár-Perl, I., Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography–mass spectrometry. *J. Chromatogr. A* **1999,** *847* (1), 91-102.

89. Hummel, J.; Strehmel, N.; Bölling, C.; Schmidt, S.; Walther, D.; Kopka, J., Mass spectral search and analysis using the golm metabolome database. In *The handbook of plant metabolomics*, Wiley-VCH Verlag GmbH & Co. KGaA: 2013; pp 321-343.

90. Kováts, E., Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: Retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. *Helv. Chim. Acta* **1958,** *41* (7), 1915-1932.

91. Wagner, C.; Sefkow, M.; Kopka, J., Construction and application of a mass spectral and retention time index database generated from plant gc/ei-tof-ms metabolite profiles. *Phytochemistry* **2003,** *62* (6), 887-900.

92. Strehmel, N.; Hummel, J.; Erban, A.; Strassburg, K.; Kopka, J., Retention index thresholds for compound matching in gc–ms metabolite profiling. *J. Chromatogr. B* **2008,** *871* (2), 182-190.

93. Lu, W.; Bennett, B. D.; Rabinowitz, J. D., Analytical strategies for lc-ms-based targeted metabolomics. *J. Chromatogr. B* **2008,** *871* (2), 236-242.

94. Cubbon, S.; Antonio, C.; Wilson, J.; Thomas-Oates, J., Metabolomic applications of hilic-lc-ms. *Mass Spectrom. Rev.* **2010,** *29* (5), 671-684.

95. Lammerhofer, M., Hilic and mixed-mode chromatography: The rising stars in separation science. *J. Sep. Sci.* **2010,** *33* (6-7), 679-680.

96. Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G., Evaluation of coupling reversed phase, aqueous normal phase, and hydrophilic interaction liquid

chromatography with orbitrap mass spectrometry for metabolomic studies of human urine. *Anal. Chem.* **2012,** *84* (4), 1994-2001.

97. Zhang, R.; Watson, D. G.; Wang, L. J.; Westrop, G. D.; Coombs, G. H.; Zhang, T., Evaluation of mobile phase characteristics on three zwitterionic columns in hydrophilic interaction liquid chromatography mode for liquid chromatography-high resolution mass spectrometry based untargeted metabolite profiling of leishmania parasites. *J. Chromatogr. A* **2014,** *1362*, 168-179.

98. Buszewski, B.; Noga, S., Hydrophilic interaction liquid chromatography (hilic)—a powerful separation technique. *Anal. Bioanal. Chem.* **2012,** *402* (1), 231-247.

99. Kloos, D. P.; Lingeman, H.; Niessen, W. M. A.; Deelder, A. M.; Giera, M.; Mayboroda, O. A., Evaluation of different column chemistries for fast urinary metabolic profiling. *J. Chromatogr. B* **2013,** *927*, 90-96.

100. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., Metlin - a metabolite mass spectral database. *Therapeutic Drug Monitoring* **2005,** *27* (6), 747-751.

101. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., Xcms: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006,** *78* (3), 779-787.

102. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G., Xcms online: A web-based platform to process untargeted metabolomic data. *Anal. Chem.* **2012,** *84* (11), 5035-5039.

103. Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G., Metlin: A technology platform for identifying knowns and unknowns. *Anal. Chem.* **2018,** *90* (5), 3156-3164.

104. Varmuza, K., Filzmoser, P. , *Introduction to multivariate statistical analysis in chemometrics*. CRC Press: Boca Raton, 2009.

105. Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B., Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **2014,** *10* (3), 361-374.

106. Larson, M. G., Analysis of variance. *Circulation* **2008,** *117* (1), 115-121.

107. Tukey, J. W., Comparing individual means in the analysis of variance. *Biometrics* **1949,** *5* (2), 99-114.

108. Levene, H., Robust tests for equality of variances. In *Contributions to probability and statistics: Essays in honor of harold hotelling*, Olkin, I., Ed. Stanford University Press: 1960; pp 278-292.

109. Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O., A guideline to univariate statistical analysis for lc/ms-based untargeted metabolomics-derived data. *Metabolites* **2012,** *2* (4), 775-795.

110. Welch, B. L., On the comparison of several mean values: An alternative approach. *Biometrika* **1951,** *38* (3/4), 330-336.

111. Games, P. A.; Howell, J. F., Pairwise multiple comparison procedures with unequal n's and/or variances: A monte carlo study. *Journal of Educational Statistics* **1976,** *1* (2), 113-125.

112. Rosenthal, J. A., Qualitative descriptors of strength of association and effect size. *J. Soc. Serv. Res.* **1996,** *21* (4), 37-59.

113. Ialongo, C., Understanding the effect size and its measures. *Biochem. Med.* **2016,** *26* (2), 150-163.

114. van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* **2006,** *7*, 142-142.

115. Marshall, D. D.; Lei, S.; Worley, B.; Huang, Y.; Garcia-Garcia, A.; Franco, R.; Dodds, E. D.; Powers, R., Combining di-esi–ms and nmr datasets for metabolic profiling. *Metabolomics* **2015,** *11* (2), 391-402.

116. Wheelock, A. M.; Wheelock, C. E., Trials and tribulations of 'omics data analysis: Assessing quality of simca-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* **2013,** *9* (11), 2589-2596.

117. Worley, B.; Powers, R., Multivariate analysis in metabolomics. *Current Metabolomics* **2013,** *1* (1), 92-107.

118. Brulle, F.; Morgan, A. J.; Cocquerelle, C.; Vandenbulcke, F., Transcriptomic underpinning of toxicant-mediated physiological function alterations in three terrestrial invertebrate taxa: A review. *Environ. Pollut.* **2010,** *158* (9), 2793-2808.

119. Ji, C. L.; Wu, H. F.; Wei, L.; Zhao, J. M.; Lu, H. J.; Yu, J. B., Proteomic and metabolomic analysis of earthworm eisenia fetida exposed to different concentrations of 2,2 ',4,4 '-tetrabromodiphenyl ether. *J. Proteomics* **2013,** *91*, 405-416.

120. Bundy, J.; Sidhu, J.; Rana, F.; Spurgeon, D.; Svendsen, C.; Wren, J.; Sturzenbaum, S.; Morgan, A. J.; Kille, P., 'Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm lumbricus rubellus. *BMC Bio.* **2008,** *6* (1), 25.

121. Edwards, C. A.; Bohlen, P. J., *Biology and ecology of earthworms*. 3 ed.; Chapman & Hall: London, UK, 1996; p 11-12.

122. Bilej, M.; Procházková, P.; Šilerová, M.; Josková, R., Earthworm immunity. In *Invertebrate immunity*, Söderhäll, K., Ed. Springer US: 2010; Vol. 708, pp 66-79.

123. Bundy, J. G.; Osborn, D.; Weeks, J. M.; Lindon, J. C.; Nicholson, J. K., An nmr-based metabonomic approach to the investigation of coelomic fluid biochemistry in earthworms under toxic stress. *FEBS Lett.* **2001,** *500* (1-2), 31-35.

124. Burch, W. S.; Fitzpatrick, C. L.; Goven, J. A.; Venables, J. B.; Giggleman, A. M., In vitro earthworm lumbricus terrestris coelomocyte assay for use in terrestrial toxicity identification evaluation. *B. Environ. Contam. Tox.* **1999,** *62* (5), 547-554.

125. Klobucar, G. I. V.; Stambuk, A.; Srut, M.; Husnjak, I.; Merkas, M.; Traven, L.; Cvetkovic, Z., Aporrectodea caliginosa, a suitable earthworm species for field based genotoxicity assessment? *Environ. Pollut.* **2011,** *159* (4), 841-849.

CHAPTER TWO

^{1}H NMR Metabolic Profiling of Earthworm (*Eisenia fetida*) Coelomic Fluid, Coelomocyte, and Tissue

Based in part on a paper published in the *Journal of Proteome Research*

J. Proteome Res. 2017, *16* (9), 3407-3417

Acknowledgments: I would like to thank Dr. Preston Williams for conducting the mass spectrometry experiments and data interpretation under the guidance of Dr. Yinsheng Wang, in addition to Drs. Luzineide Tinoco and Meredith Dinges for aiding with NMR acquisition and data interpretation.

Abstract

Earthworm metabolism is recognized as a useful tool for monitoring environmental insults and measuring ecotoxicity, yet extensive earthworm metabolic profiling using ^{1}H nuclear magnetic resonance (NMR) spectroscopy has been limited in scope. This study aimed to expand the identification of metabolites in earthworm coelomic fluid, coelomocytes and tissue to aid systems toxicology research. Fifty-nine metabolites in *Eisenia fetida* were identified, with 47 detected in coelomic fluid, 41 in coelomocytes, and 54 in whole-worm samples and tissue extracts. The newly detected but known metabolites 2-aminobutyrate, nicotinurate, *N*δ,*N*δ,*N*δ-trimethylornithine, and trigonelline are reported. We also report for the first time, a new metabolite malylglutamate. The elucidation of the malylglutamate structure is described in detail in Chapter 3, along with preliminary experiments to explore its possible biological functions.

2.1. Introduction

From relatively inert soil amendments like sulfur and lime to the transport of agricultural and industrial contaminants to foods, methods to measure and control various aspects of ecological stress are essential for crop production, water quality, and environmental health.[1-2] As described in Chapter 1, earthworms are ideal organisms for studying environmental stressors encountered in soil environments. For example, upon exposure to 3-fluoro-4-nitrophenol, Bundy, et al. [3] demonstrated that relative concentrations of trimethylamine *N*-oxide and succinate increased in the coelomic fluid (CF) of *E. veneta*, while malonate and acetate concentrations decreased. In another study, the metabolomic impact of endosulfan on the CF and whole-tissue extract of *E. fetida* was measured via the Organization for Economic Co-operation and Development (OECD) filter paper test.[4] Decreases in α-ketoglutarate, malate, spermidine, and succinate were observed in the tissue and CF of exposed worms, while alanine, ATP, betaine, lactate, and *myo*-inositol levels increased. In the whole-worm extract, fumarate also decreased in comparison to the control. Similar metabolic perturbations were found in further studies on endosulfan and endosulfan sulfate in soil.[5] Additionally, CF and whole-worm metabolomics has been used to distinguish between earthworm species, including the morphologically similar *E. fetida* and *E. andrei*.[6]

The earthworm's habitat has evolved its unique metabolome to help it adapt to changing water content and temperatures, complex food sources, and soil types. The work described in this chapter aims to provide a more complete metabolic profile, advancing our understanding of the biochemical impacts of environmental stress, and ultimately,

expanding insights into the differences in metabolism among species. In this work, nuclear magnetic resonance (NMR) spectroscopy was used to profile the *E. fetida* CF and coelomocyte (CC) metabolomes *(i.e.*, the coelom metabolome) and compare these against whole-worm and tissue extracts (*i.e.*, worm samples post-CF/CC extrusion) to better understand the biochemical role of the coelom and expand overlooked facets of the earthworm metabolome.

2.2. Experimental Procedures

2.2.1. Earthworm Culturing

Earthworms (*Eisenia fetida)* were purchased from Ward's Science (Rochester, NY, USA) and cultured in Magic Worm® Ranches containing Magic Worm® bedding (Magic Products Inc., Amherst Junction, WI, USA).[7] Earthworms were fed biweekly with Magic Worm® food and bedding was changed every four months to maintain colony health. Bedding was prepared by adding 4 L of dechlorinated water for every bag of bedding and then left for 24 h prior to transferring the earthworms. Earthworms from different ranches were mixed together when changing soils to main population homogeneity. Only adult earthworms with fully developed clitellia were used for this study.

2.2.2. Coelomic Fluid Extrusion

CF was extruded from the earthworms similarly to previously described methods.[3-4] Earthworms were dipped in EMD Millipore Simplicity® ultrapure water to remove soil and patted dry. The earthworm was transferred to a 35 x 10 mm Falcon® petri dish containing

500 μL of 0.1% NaCl in ultrapure water. A voltage was applied across the earthworm 10 times in <1 s increments using a 9V battery with a wire snap. For pooled samples, this step was repeated with additional worms. Next, the fluid was transferred to a 1.5 mL Eppendorf tube®, the petri dish was rinsed with 500 μL NaCl solution and transferred to the same Eppendorf tube. The sample was centrifuged for 20 min at 16168 x *g*, and the supernatant evaporated overnight using a Savant SC110 Speedvac® equipped with a refrigerator vapor trap (RVT400). The dried samples were stored at -80°C until analysis.

2.2.3. Coelomocyte Extraction

Following centrifugation and transfer of CF, a pellet of CC and biosolids remains. Cellular metabolism was quenched by adding 500 μL ice-cold Fischer Scientific Optima® methanol (Fair Lawn, NJ, USA) to the pellet immediately after CF transfer. The sample was vortexed and sonicated briefly, and stored on ice until extraction. An equal quantity of ice-cold ultrapure water was added to the sample and mixed using vortex and sonication. Lipids were removed with two additions of 250 μL ice-cold chloroform (Macron Fine Chemical, Center Valley, PA, USA). The aqueous layer was transferred to a fresh tube, dried via Speedvac, and stored at -80°C.

2.2.4. Whole-worm & Tissue Extraction

Metabolites were extracted from earthworms after CF/CC extrusion, referred here as tissue extracts, and whole-worm or whole organism extracts, as typically reported in literature. Earthworms were flash-frozen in liquid nitrogen, lyophilized, and homogenized

by bead-beating with 1 mL of (50:50) cold methanol:water. The sample was centrifuged at 16168 x *g* for 20 min and the supernatant was transferred to a new tube for clean-up with 500 µL ice-cold chloroform. The aqueous layer was dried and stored at -80°C.

2.2.5. Weak Cation Exchange Solid Phase Extraction

A dried 10-worm pooled CF sample was reconstituted in 2 mL ultrapure water and titrated to pH 3 with hydrochloric acid (Fisher Scientific, Fair Lawn, NJ, USA) prior to separation on a PerkinElmer Supra-Clean SPE WCX column (100 mg/3 mL) (France). The cartridge was conditioned with 6 mL of methanol and ultrapure water, loaded, and washed with 1 mL of methanol until dry. Cations were eluted with 2 mL of 84:14:2 Fischer Scientific Optima® acetonitrile:H_2O:trifluoroacetic acid (Acros, NJ, USA).

2.2.6. NMR Sample Preparation

Following lyophilization, most samples for NMR analysis were reconstituted in 500 µL of 50 mM phosphate buffer in D_2O (pD 7.45) containing 0.5 mM 3-(trimethylsilyl)propane-1-sulfonic acid-*d6* (DSS-d_6) and 0.2 mM ethylenediaminetetraacetic acid-*d16* (EDTA-d_{16}) (Cambridge Isotope Laboratories, Tewksybury, MA, USA) and vortexed until dissolved. For experiments conducted in 90% H_2O, samples were reconstituted in 500 µL of ultrapure water and vortexed until dissolved. The samples were passed through a washed 3K Amicon® Ultra 0.5 mL centrifugal filter to remove macromolecules. Centrifugal filters were washed to remove glycerol by mixing on a stir plate for ~20 hr in 400 mL ultrapure water, with the water replaced after ~5 hr. Prior

to use, the filters were rinsed a final time with 500 μL D_2O buffer and spun at 16168 x *g* for 15 min. Samples were applied to the filter and centrifuged at the same speed for 30 min. A 400 μL aliquot of the filtered extract was transferred to a 5 mm NMR tube with 200 μL buffer and mixed well prior to analysis. For the 90% H_2O samples, 60 μL of D_2O buffer was added to 400 μL of filtrated and 40 μL of ultrapure water prior to titration with deuterium chloride (Sigma Aldrich, St. Louis, MO, USA).

2.2.7. NMR Acquisition Parameters

A Bruker Avance NMR spectrometer operating at 599.52 MHz and equipped with a BBI probe and a Bruker Avance III NMR spectrometer operating at 700.23 MHz and equipped with a TCI CryoProbe were used for these experiments. ^{1}H NMR survey spectra were acquired at 600 MHz using presaturation water suppression (zgpr) with 512 scans, 32 dummy scans, 2 s delay time, 2.38 s acquisition time, 11.4864 ppm spectral width, and 32768 points at 25°C. For the ^{1}H survey spectra acquired at 700 MHz water suppression was conducted using 1D NOESY (noesypr1d) with presaturation during the 120 ms mixing time. Spectra were acquired with 256 scans, 16 dummy scans, 2 s delay time, 2.03 s acquisition time, 11.5169 ppm spectra width, and 32768 points at 25°C. At 600 MHz, the double quantum filtered COSY spectra (cosygpprqf) were measured with a 45° pulse, while the TOCSY spectra (mlevgpphw5) were measured using a mixing time of 120 ms.[8-11] Both experiments were performed with 32 scans and 16 dummy scans, with 2048 points acquired in F2 and 512 in F1. The 2D *J*-Resolved spectra (jresgpprqf) were acquired at 600 MHz with 8192 points in F2 and 64 in F1 with 64 scans and 16 dummy scans.[12] Lastly, the

^{1}H-^{13}C edited HSQCs (hsqcedetgpsisp2.2) were acquired at 700 MHz with 128 scans and 16 dummy scans into 2048 by 128 points in F2 and F1, respectively.[13-16]

2.2.8. NMR Data Processing

Bruker TopSpin 3.2 was used for initial phasing and chemical shift referencing to DSS (0 ppm) prior to processing with MestReNova 10. Free induction decays (FIDs) for ^{1}H survey spectra were apodized by multiplication with an exponential function equivalent to 1 Hz line broadening, zero-filled to 131072 points, drift corrected by 5%, and baseline corrected using Bernstein Polynomial Fit. All 2D spectra were baseline corrected using Bernstein Polynomial Fit with polynomial order 3 and the reduce T_1 noise function was applied. For the TOCSY, COSY, and HSQC, spectra were zero filled to 4096 by 2048 in both dimensions, while the 2D ^{1}H *J*-resolved spectra were zero filled to 16384 by 128. The TOCSY and HSQC spectra were apodized using a $\cos^2$ function in both dimensions, while the COSY and 2D ^{1}H *J*-resolved spectra were apodized with a $\sin^2$ function.

Authentic standards were acquired and analyzed to confirm assigned resonances unless noted with an * in Table 2.1. The Human Metabolome Database and Madison Metabolomics Consortium Database were used to aid metabolite identification.[17-18]

2.2.9. Liquid Chromatography - Tandem Mass Spectrometry Analysis

LC-MS experiments were conducted using eluate from the WCX separations. WCX eluates were first reconstituted in 100 μL ultrapure water with 0.1% trifluoracetic acid for desalting with Agilent OMIX SCX 100 μL tips, following the manufacturer's

instructions. Desalted samples were dried via Speedvac, reconstituted in 20 μL 0.1% formic acid in ultrapure water, centrifuged at 16168 x *g* for 1 min, and subjected to LC-MS.

On-line LC-MS/MS analysis was performed on an LTQ-Orbitrap Velos mass spectrometer coupled with an EASY-nLC 1000 HPLC system and a nanoelectrospray ionization source (Thermo Fisher Scientific, San Jose, CA, USA). Sample injection, enrichment, desalting, and HPLC separation were conducted automatically. The nLC was equipped with an in-house packed trapping column followed by an in-house packed separation column, both packed with ReproSil-Pur C18-AQ resin (3 μm, Dr. Maisch HPLC GmbH, Germany). The metabolites were separated using a 120-min linear gradient of 2−40% acetonitrile in 0.1% formic acid at a flow rate of 230 nL/min and electrosprayed (spray voltage 1.8 kV) into a LTQ-Orbitrap Velos mass spectrometer operated in the positive-ion mode. Full scan MS (50 – 1000 m/z) were acquired at a resolution of 60,000 (at m/z 400) and followed by data dependent acquisition (normalized collision energy of 35.0) of MS/MS for the twenty most abundant ions found in the full-scan MS exceeding a threshold of 1000 counts.

2.3. Results

This study aimed to provide a comprehensive survey of *E. fetida* endogenous metabolites using pooled and individual tissue, CF, and CC samples. ^{1}H NMR spectroscopy (Figure 2.1) was employed to detect 59 metabolites in *E. fetida*, with 47 detected in CF, 41 in CC extracts, and 54 in extracts of whole-worm samples and tissue

samples depleted of CF and CC (referred to as tissue samples). Due to the high similarity between whole-worm and tissue extracts, we will only discuss our tissue datasets and reserve discussion of whole-worm extracts to datasets reported in literature. Five metabolites: 2-aminobutyrate, malylglutamate, nicotinurate, *N*δ,*N*δ,*N*δ-trimethylornithine, and trigonelline are new to earthworm metabolomics (Table 2.1). Importantly, we believe this to be the first report of the identification of malylglutamate in any biological system. The elucidation of the structure of malylglutamate is described in detail in Chapter 3, along with preliminary experiments to explore its possible biological functions.

2.3.1. ^{1}H NMR Metabolic Profiling of Earthworm CF, CC and Tissue

Global metabolic profiling was conducted using an array of NMR experiments to determine endogenous metabolites in earthworm tissue, CF, and CC. These findings are highlighted in the ^{1}H NMR spectra in Figure 2.1 and are summarized in Tables 2.1 and 2.2. In cases where it was not possible to conclusively identify metabolite spectroscopic signatures at neutral pH, the titration of a pooled CF sample in ~0.5 unit intervals between pD 3.06 – 7.40 was useful to resolve resonances and confirm assignments (Figure 2.2). For instance, the downfield shift of malate reveals the gamma protons of glutamate (2.35 ppm, Figure 2.2B) beginning at pD 5.42 and the methyl resonance of threonine is resolved at pD 4.06, as lactate shifts downfield with increasing acidity (1.32 ppm, Figure 2.2C).

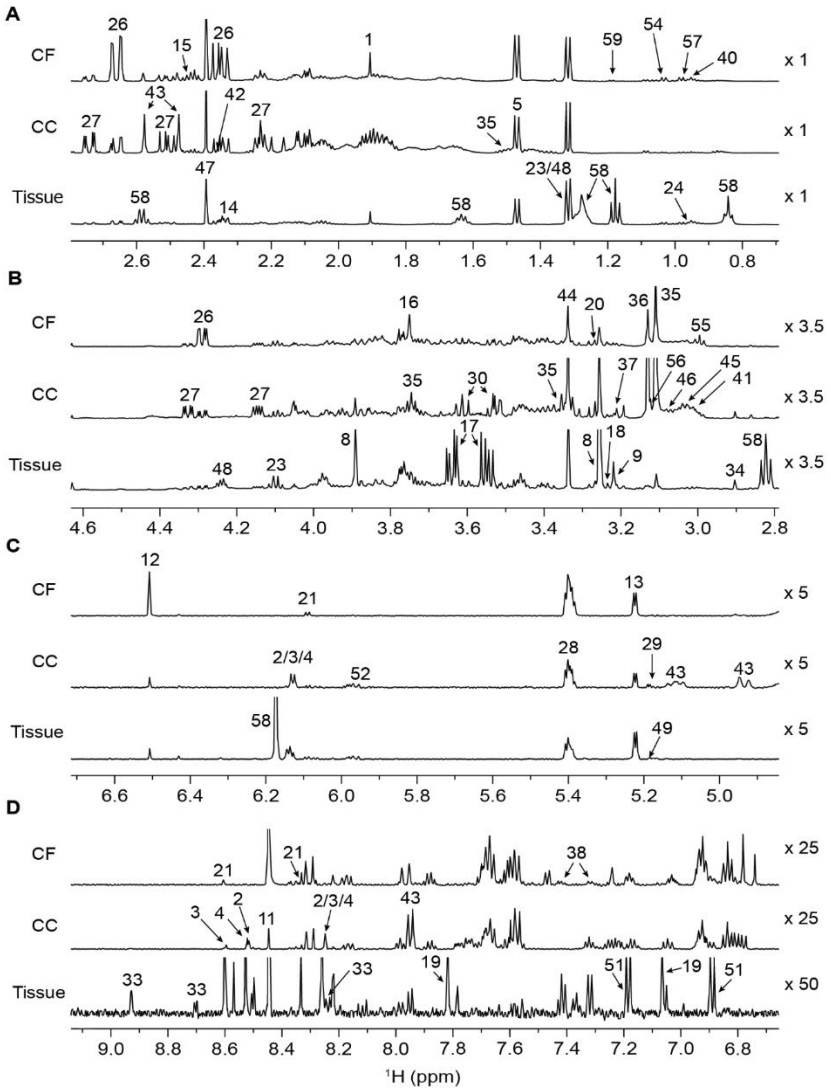


Figure 2.1. ^{1}H NMR spectra of 20-worm concentrated coelomic fluid (CF), 60-worm concentrated coelomocytes (CC), and 55 mg of pooled tissue extracts. The intensity of each spectrum is expanded is listed on the right side of each spectrum. The spectra are divided into four regions: A) 0.75-2.8 ppm; B) 2.8-4.6 ppm; C) 4.7-6.7 ppm; and D) 6.7-9.2 ppm and resonances are annotated to their corresponding metabolite number in Table 2.1. Several metabolites were identified but not annotated due to low intensity or their absence in the representative spectra.

Table 2.1. List of aqueous metabolites and their ^{1}H chemical shifts detected in pooled and individual coelomic fluid (CF), coelomocytes (CC), and tissue extracts. The NMR spectra in which metabolites are confirmed are noted for each matrix (a = 1D ^{1}H NMR, b = 2D ^{1}H *J*-Resolved, c = COSY, d = TOCSY, and e = ^{1}H-^{13}C HSQC). Bolded metabolites are new to earthworm metabolomics and the spectra of authentic standards were not recorded for metabolites marked with an *. Signal multiplicity is represented by: s= singlet, d= doublet, dd= doublet of doublets, m=multiplet, q=quartet, t= triplet.

#	Metabolite	Coelomic Fluid	Coelomocytes	Tissue	^{1}H Chemical Shifts (ppm)
1	Acetate	a,b,e	a,b	a,b,e	1.91 (s)
2	Adenosine diphosphate		a,b,c,d	a,b,c,d	4.22 (m), 4.38 (m), 4.53 (t), 6.14 (d), 8.25 (s), 8.52 (s)
3	Adenosine monophosphate		a,b,c,d	a,b,c,d	4.01 (dd), 4.36 (dd), 4.50 (dd), 6.13 (d), 8.25 (s), 8.60 (s)
4	Adenosine triphosphate		a,b,c,d	a,b,c,d	4.22 (m), 4.38 (m), 4.53 (t), 6.14 (d), 8.25 (s), 8.52 (s)
5	Alanine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.47 (d), 3.78 (q)
6	Asparagine	a,b,c,d		a,b,c,d	2.91 (m), 4.01 (dd)
7	Aspartate	a,b,c,d	a,b,c,d	a,b,c,d,e	2.74 (m), 3.89 (dd)
8	Betaine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	3.26 (s), 3.90 (s)
9	Choline	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	3.21 (s), 3.52 (m), 4.07 (m)
10	Cytosine	a,b,c,d			5.95 (d), 7.49 (d)
11	Formate	a,b,e	a,b	a,b,e	8.44 (s)
12	Fumarate	a,b,e	a,b,e	a,b,e	6.50 (s)
13	Glucose	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	3.23 (dd), 3.39 (m), 3.46 (m), 3.52 (dd), 3.73 (m), 3.82 (m), 4.83 (d), 5.22 (d)
14	Glutamate	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	2.05 (m), 2.12 (m), 2.34 (m), 3.74 (q)
15	Glutamine	a,b,c,d,e	a,b,c,d,e	a,b,c,d	2.02 (m), 2.12 (m), 3.76 (t)
16	Glycine	a,b,c,d	a,b,c,d,e	a,b,c,d,e	3.54 (s)
17	Glycerol			a,b,c,d,e	3.60 (m), 3.77 (m)
18	Glycerophosphocholine	a,b,e	a,b,e	a,b,e	3.22 (s), 3.64 (m), 3.91 (m), 4.32 (m)
19	Histidine	a,b,c,d		a,b,c,d	3.13 (dd), 2.23 (dd), 3.97 (dd), 7.06 (d), 7.80 (d)
20	Histidine-betaine	a,b,e	a,b,e	a,e	3.26 (s)
21	Inosine	a,b,c,d	a,b,c,d	a,b,c,d	3.83 (dd), 3.91 (dd), 4.27 (dd), 4.43 (dd), 6.09 (d), 8.23 (s), 8.33 (s)
22	Isoleucine	a,b,c,d	a,b,c,d	a,b,c,d,e	0.93 (t), 1.00 (d), 1.25 (m), 1.46 (m), 1.97 (m), 3.66 (d)
23	Lactate	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.32 (d), 4.10 (q)
24	Leucine	a,b,c,d	a,b,c,d	a,b,c,d,e	0.95 (t), 1.71 (m), 3.72 (m)
25	Lysine			a,b,c,d,e	1.47 (m), 1.72 (m), 1.90 (m), 3.02 (t), 3.74 (t)
26	Malate	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	2.36 (dd), 2.66 (dd), 4.29 (dd)
27	**Malylglutamate***	**a,b,c,d,e**	**a,b,c,d,e**	**a,b,c,d,e**	**1.88 (m), 2.05 (m), 2.23 (m), 4.14 (dd), 2.52 (dd), 2.74 (dd), 4.33 (dd)**

28	Maltose-x	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	3.27 (dd), 3.41 (t), 3.58 (m), 3.63 (m), 3.66 (m), 3.70 (m), 3.76 (m), 3.84 (m), 3.90 (dd), 3.93 (d), 3.96 (m), 5.22 (d), 5.40 (d)
29	Mannose			a,b,c,d	3.37 (ddd), 3.56 (t), 3.65 (m), 3.74 (m), 3.80 (m), 3.84 (m), 3.88 (dd), 3.93 (m), 5.17 (m)
30	*myo*-Inositol	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	2.67 (t), 3.52 (dd), 3.61 (t), 4.05 (t)
31	Nicotinamide			a,b,c,d	7.57 (m), 8.22 (m), 8.69 (dd), 8.90 (dd)
32	Nicotinate			a,b,c,d	7.51 (dd), 8.24 (m), 8.60 (dd), 8.93 (d)
33	**Nicotinurate**			**a,b,c,d**	**3.96 (s), 7.57 (m), 8.23 (tt), 8.68 (dd), 8.91 (m)**
34	*N*,*N*-Dimethylhistidine	a,b,c,d	a,b,c,d	a,b,c,d	2.90 (s), 3.25 (dd), 3.86 (dd), 7.04 (s), 7.76 (s)
35	*N*ε,*N*ε,*N*ε-Trimethyllysine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.46 (m), 1.89 (m), 3.11 (s), 3.34 (m), 3.75 (t)
36	***N*δ,*N*δ,*N*δ-Trimethylornithine***	**a,b,c,d,e**	**a,b,c,d,e**	**a,b,c,d,e**	**1.94 (m), 3.13 (s), 3.38 (m), 3.78 (t)**
37	*N*,*N*,*N*-Trimethyltaurine*	a,b,e	a,b,e	a,b,e	3.19 (s)
38	Phenylalanine	a,b,c,d		a,b,c,d,e	3.20 (m), 3.98 (dd), 7.32 (d), 7.36 (m), 7.42 (m)
39	Proline-betaine			a,b,e	3.10 (s), 3.29 (s)
40	Propionate	a,b,c,d		a,b,c,d	1.05 (t), 2.17 (m)
41	Putrescine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.76 (m), 3.05 (m)
42	Pyruvate	a,b,e	a,b,e	a,b,e	2.35 (s)
43	Riboflavin	a,b,c,d,e	a,b,c,d,e	a,b,c,d	2.47 (s), 2.58 (s), 3.72 (dd), 3.87 (dd), 3.92 (m), 3.97 (m), 4.43 (m), 4.95 (m), 5.13 (m), 7.95 (s), 7.97 (s)
44	*scyllo*-Inositol	a,b,e	a,b,e	a,b,e	3.33 (s)
45	Spermidine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.78 (m), 2.11 (m), 3.08 (m)
46	Spermine	a,b,c,d,e	a,b,c,d,e	a,b,c,d,e	1.82 (m), 2.12 (m), 3.11 (m)
47	Succinate	a,b,e	a,b,e	a,b,e	2.40 (s)
48	Threonine	a,b,c,d	a,b,c,d,e	a,b,c,d,e	1.32 (d), 3.57 (d), 4.24 (m)
49	Trehalose	a,b,c,d	a,b,c,d	a,b,c,d	3.44 (t), 3.64 (dd), 3.76 (m), 3.81 (m), 5.18 (d)
50	**Trigonelline**			**a,b,c,d**	**4.43 (s), 8.08 (m), 8.83 (m), 9.11 (s)**
51	Tyrosine	a,b,c,d		a,b,c,d,e	3.05 (dd), 3.19 (dd), 3.93 (dd), 6.89 (m), 7.18 (m)
52	Uridine 5'-diphospho-*N*-acetylglucosamine	a,b,c,d	a,b,c,d	a,b,c,d	2.07 (s), 3.55 (dd), 3.80 (m), 3.86 (dd), 3.92 (dd), 3.98 (tt), 4.21 (m), 4.28 (m), 4.36 (m), 5.51 (dd), 3.96 (d), 5.97 (d), 7.95 (d)
53	Uridine	a,b,c,d			3.79 (dd), 3.90 (dd), 4.12 (m), 4.22 (dd), 4.34 (dd), 5.89 (d), 5.91 (d), 7.86 (d)
54	Valine	a,b,c,d	a,b,c,d	a,b,c,d,e	0.98 (d), 1.03 (d), 2.27 (m), 3.60 (d)
55	α-Ketoglutarate	a,b,c,d,e	a,b,c,d	a,b,c,d	2.43 (t), 3.00 (t)
56	γ-Butyrobetaine	a	a		3.02 (m), 2.26 (t), 3.12 (s), 3.31 (m)
57	**2-Aminobutyrate**	**a,b,c,d**			**0.97 (t), 1.89 (m), 3.70 (t)**
58	2-Hexyl-5-ethyl-furan-3-sulfonate (HEFS)*			a,b,c,d,e	0.84 (t), 1.17 (t), 1.28 (m), 1.63 (m), 2.58 (q), 2.82 (t), 6.17 (s)
59	3-Hydroxybutyrate	a,b,c,d			1.18 (d), 2.29 (m), 2.39 (m), 4.13 (m)

Because of the complexity of the earthworm samples, correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) were employed to reveal coupled spins and spin systems, assisting the identification of metabolites such as the heavily overlapped polyamines: putrescine, spermidine, and spermine, as annotated in the representative COSY spectrum in Figure 2.3.[8-11] Further confidence in resonance assignments was gained using homonuclear 2D *J*-resolved spectra to differentiate overlapping peaks. These spectra were was particularly useful for assigning singlets that were filtered in the double quantum filtered COSY spectra or were uncoupled to any spin system in the TOCSY spectra.[12] The 2D ^{1}H *J*-resolved spectra greatly aided the assignment of the betaine analog singlets between 3.0-3.4 ppm (Figure 2.4A), including betaine, choline, glycerophosphocholine, histidine-betaine, *N*ε,*N*ε,*N*ε-trimethyllysine, *N*,*N*,*N*-trimethyltaurine, and proline-betaine. Measurement of ^{13}C chemical shifts using the ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) experiment at 700 MHz was valuable for the validation of resonance assignments (Table 2.2), as demonstrated in the $N(CH_3)_3$ region (Figure 2.4B) in which the ^{1}H-^{13}C chemical shift pairs affirmed the resonance assignments of betaine analogs.[13-16] With these several layers of metabolite confirmation, compounds identified by NMR are well-vetted, thus providing high levels of confidence in their assignments.

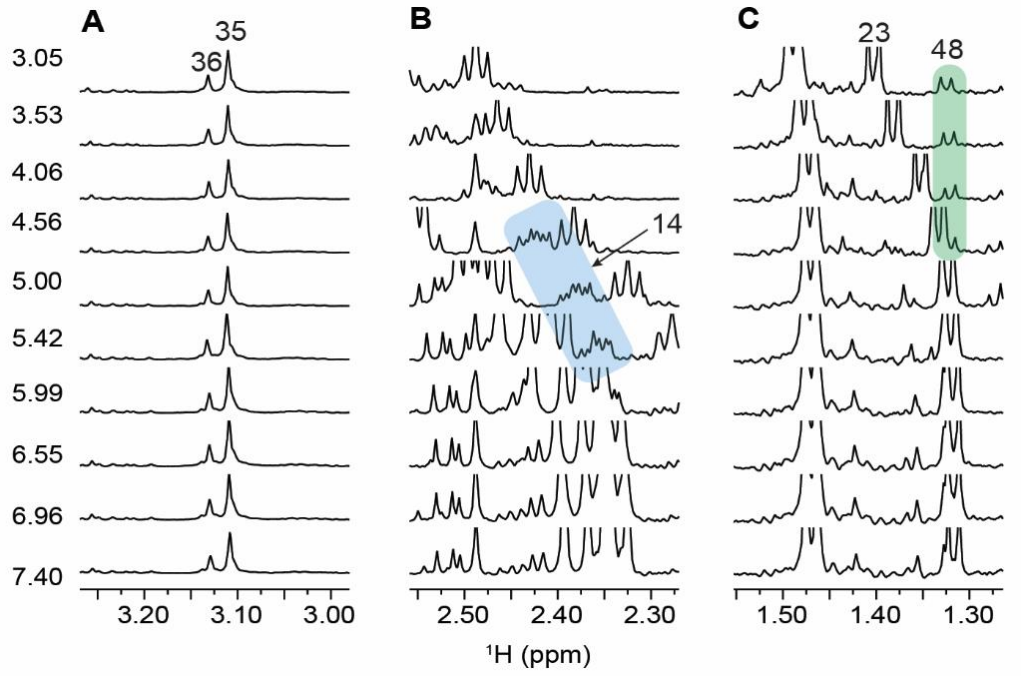


Figure 2.2. ^{1}H NMR spectra of concentrated 20-worm coelomic fluid sample that was divided into ten samples and titrated to the pD labeled on the left of each spectrum. A) the $N(CH_3)_3$ region between 3.0-3.2 ppm; B) the glutamate (14) gamma proton is revealed as the pD decreases between 2.2-2.5 ppm; and C) the methyl resonance of threonine (48) appears as lactate shifts downfield with increasing acidity. Metabolite numbering corresponds with Table 2.1 and 2.2.

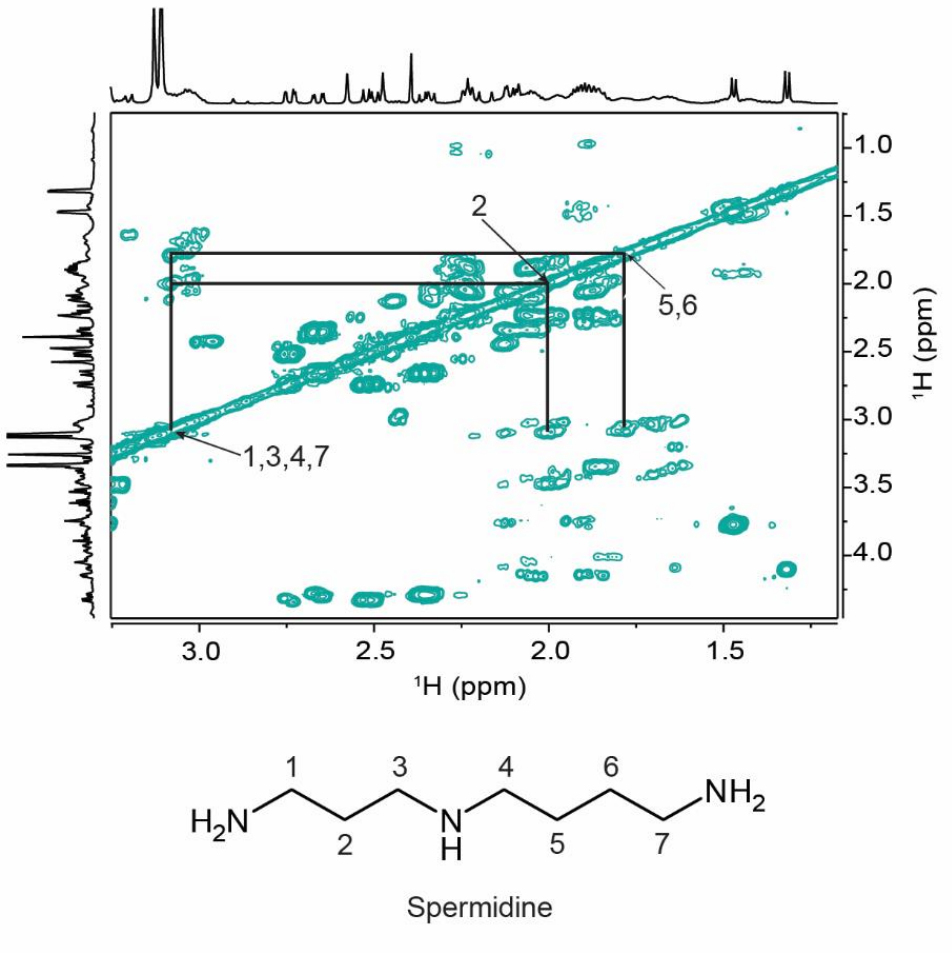


Figure 2.3. Representative COSY spectrum showing the identification of spermidine in a concentrated CF sample. Cross peak labels correspond to the proton's number in the spermidine chemical structure and lines show how resonances are coupled with each other.

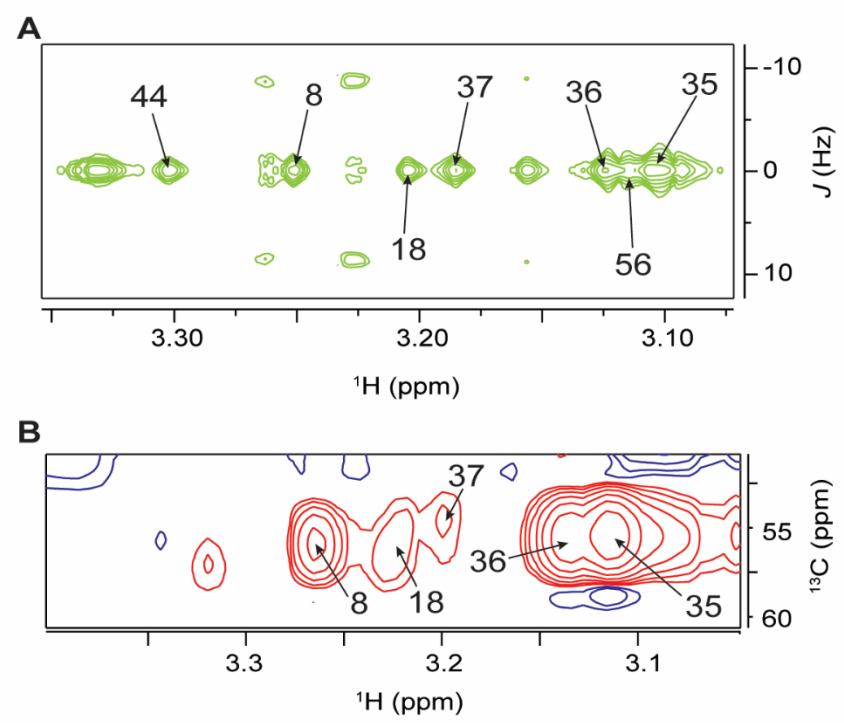


Figure 2.4. Portions of representative annotated homonuclear and heteronuclear 2D spectra demonstrating their complementary use in assigning the resonances of betaine analogs. A) The homonuclear 2D *J*-resolved spectrum showing the chemical shift (ppm) and *J*-coupling (0 Hz) used to assign singlet resonances which were overlapped with coupled resonances in the 1D spectrum. Depicted is the $N(CH_3)_3$ region of the detected betaine analogs in a concentrated CF sample. The assignment of these resonances was further vetted with B) the ^{1}H-^{13}C HSQC spectrum. The multiplicity-edited HSQC experiment allows for antiphasing of CH_2 (blue) and CH or CH_3 (red) resonances facilitating resonance identification. The chemical shifts of the $[N(CH_3)_3]$ protons of the betaine analogs appear red in the spectrum, confirming the presence of a CH_3 or CH group. The ^{1}H-^{13}C chemical shift pairs were compared and assigned based on literature data.[19] Metabolite numbering corresponds to that of Tables 2.1 and 2.2.

Table 2.2. List of aqueous metabolites and their ^{13}C chemical shifts detected using ^{1}H-^{13}C HSQC at 700 MHz. Experiments were performed using a 20-worm concentrated coelomic fluid (CF) sample, a 60-worm concentrated coelomocyte (CC) extract, and a pooled 50 mg tissue extract. Bolded metabolites are new to earthworm metabolomics and numbering corresponds to Table 2.1. All chemical shifts are from Human Metabolome Database, unless noted with a [1] to indicate shifts from the Madison Metabolomics Consortium Database, [2] to represent shifts reported by Liebeke and Bundy [19], or [3] shifts taken directly from the measured spectra.[28-29]

#	Metabolite	Coelomic Fluid	Coelomocytes	Tissue	^{13}C Chemical Shifts (ppm)
1	Acetate	x		x	26.09
2	Adenosine triphosphate			x	67.10, 67.79, 72.96, 76.96, 86.61, 89.28, 142.5, 155.37
5	Alanine	x	x	x	19.03, 53.56
7	Aspartate			x	39.33, 39.48, 55.09
8	Betaine	x	x	x	55.86, 68.64
9	Choline	x	x	x	58.49, 70.15, 56.70
11	Formate	x		x	172.41
12	Fumarate	x	x	x	138
13	Glucose	x	x	x	63.35, 63.41, 63.47, 72.34, 74.14, 74.20, 75.64, 76.96, 78.57, 94.94, 98.71
14	Glutamate	x	x	x	29.82, 36.36, 57.64
15	Glutamine	x	x		29.29, 33.93, 57.23
16	Glycine		x	x	44.3
17	Glycerol			x	65.40, 65.49, 74.98
18	Glycerophosphocholine[2]	x	x	x	56.6
20	Histidine-betaine[2]	x	x	x	55.2
22	Isoleucine			x	13.91, 17.36, 26.90, 38.69, 62.52
23	Lactate	x	x	x	22.90, 71.37
24	Leucine			x	23.57, 24.8, 26.78, 42.60, 56.21
25	Lysine			x	24.04, 29.15, 32.65, 57.45
26	Malate	x	x	x	45.46, 45.45, 73.24
27	**Malylglutamate[3]**	**x**	**x**	**x**	**31.35, 37.05, 43.96, 58.16, 72.41**
28	Maltose-x	x	x	x	62.96, 63.16, 63.29, 71.81, 72.46, 73.70, 74.02, 75.32, 75.58, 75.78, 76.62, 76.95, 78.77, 79.29, 94.51, 98.41, 102.06, 102.55
30	*myo*-Inositol	x	x	x	73.96, 74.93, 75.13, 77.16
35	*N*ε,*N*ε,*N*ε-Trimethyllysine[3]	x	x	x	24.23, 24.81, 32.68, 55.61, 57.15, 68.81

36	***Nδ, Nδ, Nδ*-Trimethylornithine[3]**	**x**	**x**	**x**	**28.53, 55.61, 56.73, 68.18**
37	*N*,*N*,*N*-Trimethyltaurine[2]	x	x	x	54.59
38	Phenylalanine			x	39.15, 58.93, 130.43, 131.81, 132.12
39	Proline-betaine			x	48.1, 54.8
41	Putrescine	x	x	x	26.73, 41.76
42	Pyruvate	x	x	x	29.23
43	Riboflavin	x	x		21.38, 23.34, 49.80, 49.85, 65.65, 65.75, 72.68, 75.28, 76.05, 118.68, 132.22
44	*scyllo*-Inositol[1]	x	x	x	76.29
45	Spermidine	x	x	x	25.21, 26.56, 26.64, 39.24, 41.19, 46.96, 49.6
46	Spermine	x	x	x	25.55, 26.58, 39.50, 47.43, 49.81
47	Succinate	x	x	x	36.83
48	Threonine		x	x	22.30, 63.46, 68.91
51	Tyrosine			x	38.27, 58.99, 118.89, 133.49
54	Valine			x	19.41, 20.75, 31.89, 63.35
55	α-Ketoglutarate	x			33.42, 38.63
58	2-Hexyl-5-ethyl-furan-3-sulfonate (HEFS)[3]			x	14.06, 16.21, 23.22, 24.75, 28.63, 30.43, 30.71, 33.55, 105.65

2.3.2. Comparison of the CF, CC and Tissue Metabolomes

Thirty-three common metabolites were observed by NMR in CF, CC, and tissue extracts, mostly consisting of sugars, organic acids, and amino acids (Table 2.1). 2-Aminobutyrate, asparagine, cytosine, histidine, 3-hydroxybutyrate, phenylalanine, propionate, and uridine were only detected in the CF and tissue, while ADP, AMP, and ATP were only found in the CC and tissue extracts. ATP was previously reported in CF but was not detected in our samples.[4] TCA cycle constituents comprise a major component of the coelom metabolome and their perturbation can give insights into bioenergetic stress. The ^{1}H NMR spectra of tissue extracts also contain the resonances of many of these constituents, though small changes in the levels of these metabolites may be more difficult to discern because of high resonance overlap. Tissue extracts contain mostly amino acids, while except for alanine the amino acids tend to be minor components of the coelom.

Recently, Liebeke, et al. [20] unlocked a key defense required for biogeochemical cycling: dialkylfuransulfonate (termed drilodefensin) metabolites, which protect earthworms from the multitude of plant polyphenols to which they are exposed through their diet. Tissue extracts contain a substantial concentration of the drilodefensin 2-hexyl-5-ethyl-furan-3-sulfonate (HEFS) (Figure 2.1) which is not observed in the coelom, in addition to glycerol, lysine, mannose, nicotinamide, nicotinate, nicotinurate, and trigonelline (Tables 2.1 and 2.2).

The high degree of similarity between the ^{1}H NMR spectra of CC and CF suggests that there is a close interaction between the two compartments. As observed in Figure 2.1, the lower intensity of the sugar resonances in the coelomocytes reveals additional

resonances from *myo*-inositol, *N*ε,*N*ε,*N*ε-trimethyllysine (TML), and riboflavin. The resonances of glutamine and glutamate are also less overlapped with other signals in the coelomocyte spectra. Metabolite investigations in CF, CC, and tissue provide complementary and unique insights into earthworm metabolism.

2.3.3. Identification of *N*δ,*N*δ,*N*δ-Trimethylornithine – a New Metabolite in CF

*N*δ,*N*δ,*N*δ-Trimethylornithine or TMO (3.129 ppm) [CAS: 66101-16-4] was identified as the neighboring singlet to *N*ε,*N*ε,*N*ε-trimethyllysine or TML (3.108 ppm), as shown in Figure 2.5A, using a combination of NMR and high resolution mass spectrometry (HRMS). In the CF pH titration experiments, the chemical shifts of $N(CH_3)_3$ groups were affected minimally with increasing acidity (Figure 2.2A), demonstrating that the singlets belonged to non-titratable groups. A weak cation exchange (WCX) solid-phase extraction strategy was employed simplify the mixture but did not offer any additional NMR insights due to the high chemical shift similarity of $N(CH_3)_3$ groups (Figure 2.6A). To aid identification, the WCX eluate was subjected to LC-MS, allowing detection of the TMO parent ion at *m/z* 175.1431 (the calculated *m/z* is 175.1463 for the $[M]^+$ ion; Figure 2.6B) and ^{13}C isotopes. Figure 2.5C presents the MS/MS spectrum of the $[M+H]^+$ ion, yielding major fragments at *m/z* 116.1 $[M - N(CH_3)_3]^+$ and 60.1 $[HN(CH_3)_3]^+$, and minor fragment ions at *m/z* 157.0 $[M - H_2O]^+$, 147.1 $[M - CO]^+$; 133.1 $[M - CO_2]^+$; and 130.1 $[M - HN(CH_3)_2]^+$. In this context, the loss of dimethylamine (*i.e.*, $HN(CH_3)_2$) may be initiated by the migration of a methyl group from the trimethylammoium moiety, and such migration was observed previously for singly charged peptides harboring a

trimethyllysine.[21] The fragmentation of TMO, TML, and betaine (Figure 2.6B-E), is consistent with reported fragmentation of betaine analogs.[22]

The chemical shift of the TMO N$\delta(CH_3)_3$ protons is downfield of the TML resonance, consistent with the additional CH_2 group of TML. Also, the resonance is consistent with the reported N$\delta(CH_3)_3$ shift of ornithine betaine, which has ornithine trimethylated on both amines.[19] The other TMO resonances are observed in our TOCSY spectrum (Figure 2.5D) at similar chemical shifts to TMO as described in Patti, et al. [23]. Finally, the TMO:TML ratios were calculated as 1:5 and 1:3, respectively, from the HRMS and NMR results. It is expected that these ratios are not identical due to differences in ionization, resolution, and sample preparation; however, the similarity of the two ratios obtained with orthogonal analytical platforms further supports the assignment of TMO as the unknown.

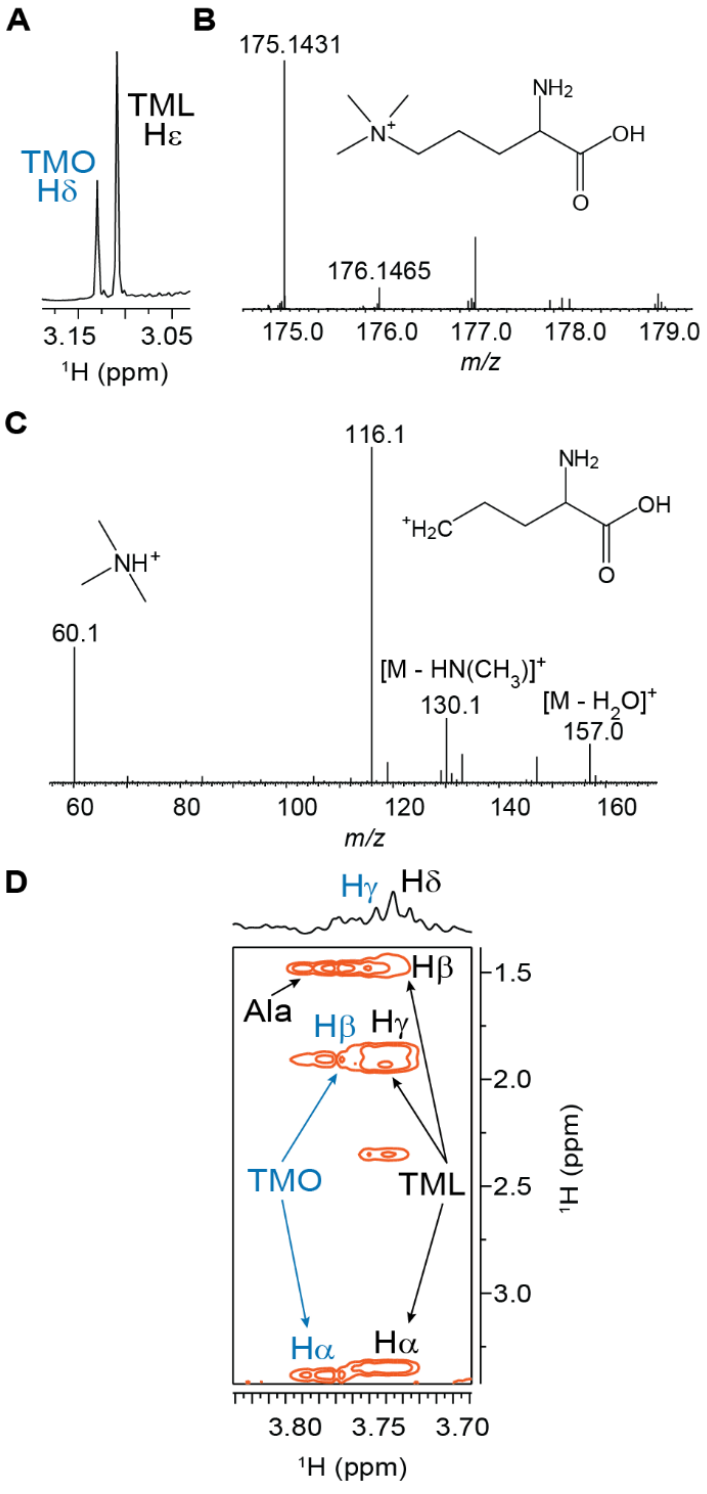

A
TML
Hε
TMO
Hδ
3.15
3.05
¹H (ppm)
B
175.1431
NH2
N+
OH
O
176.1465
175.0
176.0
177.0
178.0
179.0
m/z
C
116.1
NH2
OH
+H2C
O
NH+
60.1
[M - HN(CH3)]+
130.1
[M - H2O]+
157.0
60
80
100
120
140
160
m/z
D
Hγ
Hδ
Hβ
Ala
Hβ
Hγ
TMO
TML
Hα
Hα
1.5
2.0
2.5
3.0
¹H (ppm)
3.80
3.75
3.70
¹H (ppm)

Figure 2.5. LC-MS identification of TMO and NMR confirmation. A) The $N(CH_3)_3$ singlets of TML and TMO resonate between 3.05-3.15 ppm in the 1D NMR spectrum; B) Positive-ion ESI-MS of TMO (the calculated *m/z* 175.1464 for the M^+ ion), shown are the monoisotopic peak at *m/z* 176.1431 and the structure of TMO; C) MS/MS for the M^+ ion of TMO; and D) CF TOCSY spectrum with labeled TMO and TML spin systems. The TMO $N\delta(CH_3)_3$ and TML $N\varepsilon(CH_3)_3$ singlets are not correlated with these spin systems because the carbon-bound protons on either side of the nitrogen are too distant to be coupled. Overlapped resonances of TML and alanine (Ala) are annotated to emphasize that the peak does not belong to the TMO spin system.

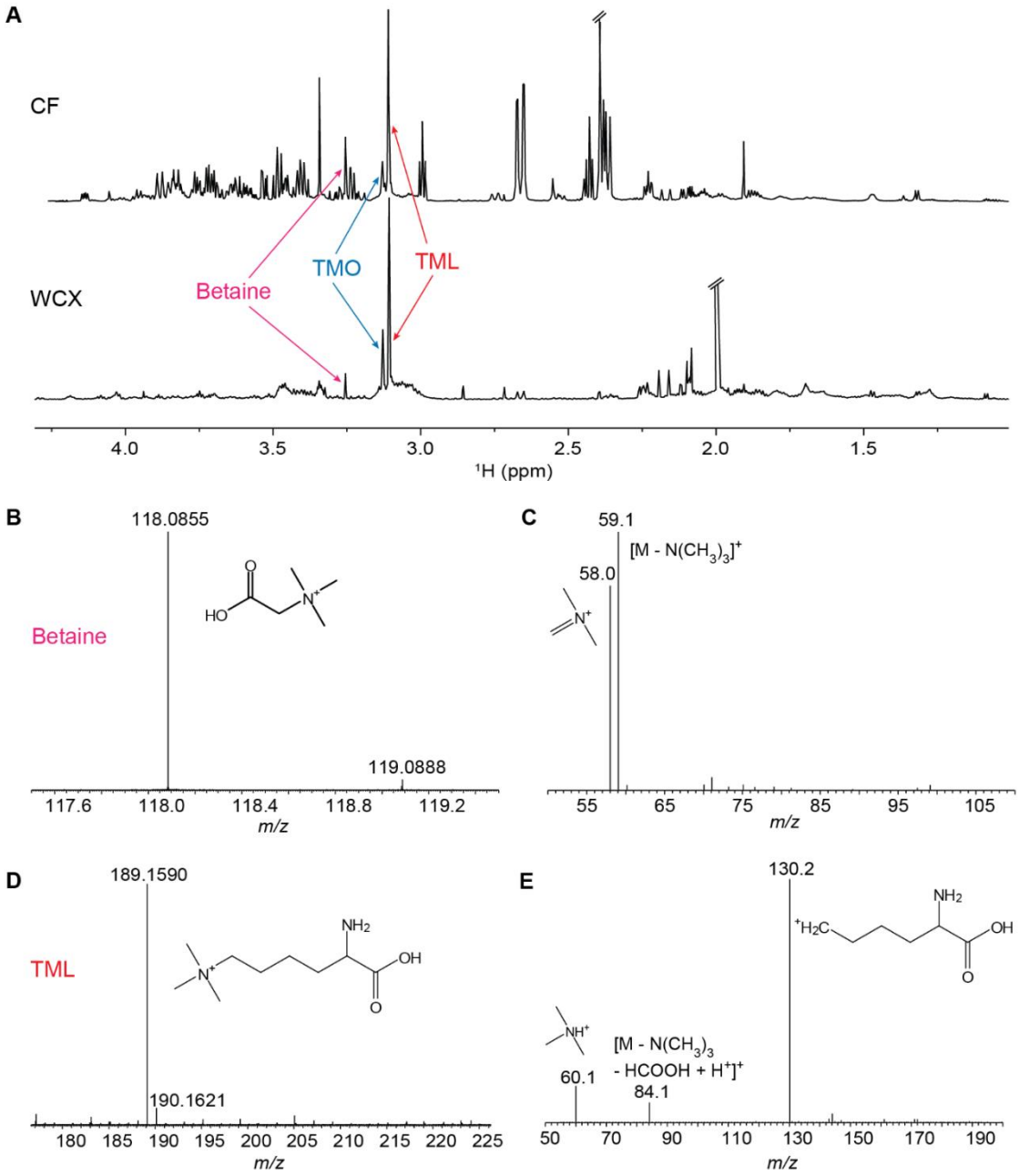

A
CF
WCX
Betaine
TMO
TML
4.0
3.5
3.0
2.5
2.0
1.5
¹H (ppm)
B
118.0855
Betaine
HO
O
N+
119.0888
117.6
118.0
118.4
118.8
119.2
m/z
C
59.1
[M - N(CH3)3]+
58.0
N+
55
65
75
85
95
105
m/z
D
189.1590
TML
NH2
OH
O
N+
190.1621
180
185
190
195
200
205
210
215
220
225
m/z
E
130.2
NH2
+H2C
OH
O
NH+
[M - N(CH3)3
- HCOOH + H+]+
60.1
84.1
50
70
90
110
130
150
170
190
m/z

Figure 2.6. Weak cation exchange (WCX) SPE of cationic metabolites in CF and LC-MS identification of known metabolites. A) ^{1}H NMR spectra comparing CF before and after WCX, showing the isolation of betaine analogs. The WCX ^{1}H NMR spectrum did not offer any additional insights into TMO's structure, which led us to turn to LC-MS to identify the compound giving rise to this resonance. Other betaine analogs detected in our LC-MS data are highlighted in B-E. B) Positive-ion ESI-MS of betaine (the calculated *m/z* 118.0868 for M^+ ion), isotopic peak at *m/z* 119.0847, and structure; C) annotated MS/MS for the M^+ ion; D) Positive-ion ESI-MS of *N*ε,*N*ε,*N*ε-trimethyllysine (the calculated *m/z* 189.1603 for the M^+ ion), isotopic peak at *m/z* 190.162, and structure; and E) annotated MS/MS for the M^+ ion.

2.3.4. Other Newly Identified Metabolites in CF, CC and Tissue

2-Aminobutyrate in the CF, and nicotinurate and trigonelline in tissue extracts were also newly detected metabolites assigned using 1D and 2D NMR spectra and confirmed by comparison to authentic standards (Figure 2.7). Previously maltose has been reported in earthworm extracts, but here we report the resonances as maltose-x due to the several overlapping resonances of maltose, maltotriose, and potentially other closely related sugars.

Rochfort, et al. [24] recently described *E. fetida* and *E. andrei* specific aromatic metabolites that we suspect may correlate with resonances in our CF NMR spectra and peaks detected in our DI-MS experiments, but due to the limited HRMS and NMR data reported[24] and the lack of commercially available standards, we could not confidently assign the MS ion or similar resonances as the reported metabolites. Earthworms contain an abundance of unique metabolites, and further elucidation of new metabolites and probing of their biochemical function could aid environmental monitoring and diagnostics.

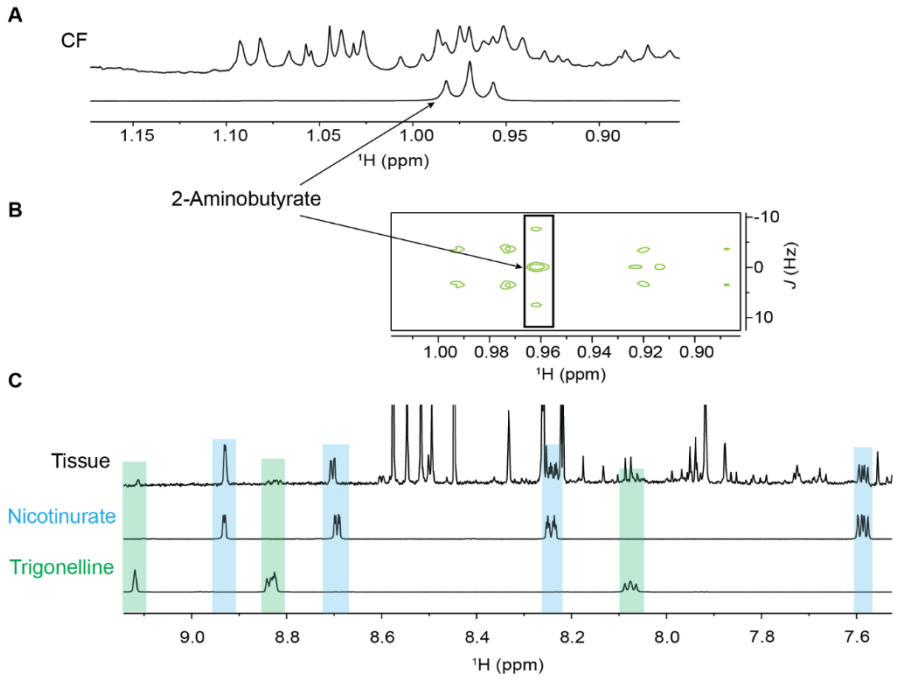


Figure 2.7. Depiction of the use of A) 1D and B) homonuclear 2D *J*-resolved spectroscopy to detect 2-aminobutyrate in the heavily overlapped methyl region of a concentrated CF sample. C) Stacked spectra showing the identification of nicotinurate (blue) and trigonelline (green) in earthworm tissue extracts.

2.4. Discussion

In this work, the complex and metabolite-rich *E. fetida* CF, CC, and tissue metabolomes were characterized using NMR spectroscopy. The abundance of TCA metabolites, osmolytes, polyamines, and other metabolites detected in these studies indicates that the CF may serve as a metabolic reservoir. Our elucidation of the coelomocyte metabolome offers a new source to evaluate earthworm health, and possibly provide insights into hepatocytic and immune function, while tissue extracts offer a

metabolic summary of global processes occurring within the earthworm. Notably, this study is the first to report 2-aminobutyrate, nicotinurate, *N*δ,*N*δ,*N*δ-trimethylornithine, and trigonelline in earthworm samples.

2.4.1. Betaine Analogs and Osmoregulation

As the cationic counterpart to malylglutamate and other organic acids, earthworms contain an arsenal of betaine analogs, including: betaine, choline, glycerophosphocholine, histidine-betaine, *N*,*N*-dimethylhistidine, *N*,*N*,*N*-trimethyltaurine, TML, and TMO. Proline-betaine was only detected in tissue samples, while γ-butyrobetaine was only detected in the CC and CF. Liebeke and Bundy [19] reported γ-butyrobetaine in the ^{1}H-^{13}C HSQC spectra measured at 800 MHz of *E. fetida* whole-worm extracts, potentially due to increased sensitivity compared to our HSQC spectrum measured at 700 MHz. These authors also detected carnitine and hydroxyproline-betaine in the whole-worm extracts of other earthworm species,[19] but these were not detected in our NMR spectra. The function of betaine analogs is unclear, but it is hypothesized that they may serve as osmoprotective agents for proteins and membranes. Earthworms must constantly adjust to changing moisture content in soil, and CF contains many other well-known osmolytes like betaine, glucose, inositol, trehalose, and polyamines. For instance, upregulation of glucose and alanine was observed in *Enchytraeus albidus* under drought conditions, while only alanine level increased in *A. caliginosa* in response to estivation.[25-26] Future studies should explore the role of betaines under similar contexts, and aim to establish the biochemical functions of this diverse group of earthworm metabolites.

It should also be noted that TML, γ-butyrobetaine, α-ketoglutarate, and succinate are a part of the carnitine biosynthesis pathway, which is essential for converting fatty acids into energy via mitochondrial and peroxisomal β-oxidation. TMO was recently revealed as a suitable substrate for TML hydroxylase, converting 75% of TMO to 3-hydroxy-*N*δ,*N*δ,*N*δ-trimethylornithine.[27] This may suggest that TMO and other betaine analogs may act as an alternative substrate when TML is limited. In mammals, the liver is the central location of the final conversion of γ-butyrobetaine by γ-butyrobetaine dioxygenase to carnitine.[28] Since γ-butyrobetaine is only observed in the coelom and some subsets of coelomocytes serve liver-like functions, this may explain the presence of γ-butyrobetaine in CC extracts.

2.4.2. Polyamines

E. fetida also contains a significant quantity of polyamines (putrescine, spermidine, and spermine) within its CC and CF. Polyamines are fully protonated under physiological conditions consistent with their various functions, which include moderating ion channels, cell growth and differentiation, gene expression, apoptosis, and chromatin status.[29] Disruption of polyamine synthesis, indicated by their downregulation, has been suggested as a sign of genotoxicity, while their upregulation acts as a protective response against oxidative, osmotic, and thermal stress. Consistent with their observation in *E. fetida,* organisms that are able to regenerate have a close association with polyamine metabolism and synthesis.[30] Putrescine was upregulated in response to removal and regeneration of worm heads. In the same study, osmotic and heat shock temporarily upregulated putrescine

and spermidine. Elevation in putrescine was additionally observed in *L. rubellus* exposed to high levels of nickel.[31] Similarly, upregulation of putrescine and alanine in two *A. caliginosa* populations exposed to the fungicide epoxyiconazole was credited as a stress indicator.[32] Moreover, endosulfan and endosulfan sulfate exposure was hypothesized to induce genotoxicity that resulted in apoptosis and thus, subsequently decreased spermidine concentrations in *E. fetida* CF.[4-5] This suggests that the CF may act as a reservoir for polyamines and probing their alteration may prove to be a useful indicator of environmental stress.

2.4.3. Drilodefensins

Some distinctive earthworm metabolites were not observed in NMR spectra of coelomocytes or CF, like the drilodefensin metabolite 2-hexyl-5-ethyl-furan or HEFS, which has been indicated as a potential biomarker for exposure to many compounds like perfluorooctane sulfonate (PFOS) and 2,2',4,4'-tetrabromodiphenyl ether (BDE-47).[20, 33-34] Liebeke *et al.* (2015) recently evaluated the plethora of drilodefensin metabolites present in the earthworm and elucidated their surfactant-like properties that help reduce protein solubilization in the presence of plant polyphenols. Drildofensins are primarily present in the earthworm gut and are upregulated under high polyphenol diets.[20] This may explain the absence of these compounds from the coelomic fluid, but investigation of the fluid under conditions that favor drildofensin upregulation would need to be conducted to confirm this finding.

2.4.4. Coelomocytes

Unlike vertebrates, invertebrates do not contain an adaptive immune system, thus a robust and diverse innate immune system is essential to combat pathogens and tend wounds. Coelomocytes and coelomic fluid of earthworms have been noted to play a central role in immunity. Worms must defend against microbial infiltration through the dorsal pores that connect the coelom with the outside environment. Proteins in the coelomic fluid have been demonstrated to hold mitogenic activity and include lysein, a protein that targets sphingomyelin.[35-36] Furthermore, antimicrobial peptides have been discovered in earthworms as well as enzymes with antioxidant capabilities, including catalase and glutathione-S-transferase.[37-38] [See Bilej *et al.* (2010) for a thorough review of earthworm immunology.][39] Together, these factors contribute to coelomic fluid's antimicrobial, hemolytic, cytotoxic, and proteolytic properties and demonstrate the potential to use coelomocytes and coelomic fluid to probe the impact of genotoxic and immunotoxic substances.

Giving its characteristic bright-yellow hue, riboflavin is found at substantial concentrations within the CC and CF. Riboflavin stored in CC has been identified as a coelomocyte chemoattractant in six earthworm species and may serve as a recruiting factor to combat microbial invasion.[40-41] Interestingly, riboflavin content in CC has also been explored as an indicator of heavy metal toxicity and soil health. Plytycz, et al. [42] hypothesized that the heavy metal body burden is localized in the chloragogenous tissue of the worm, a source of a type of CC called elocytes, explaining how heavy metal exposure interferes with worm's immune response. In this study, stimulation of the immune system

by chronic heavy metal exposure led to riboflavin depletion in the CC of *Dendrodrilus rubidus*, marking the potential of riboflavin in elocytes as an indicator of heavy metal toxicity.[41] These studies highlight the potential of riboflavin levels in coelomocytes as a potential signature of metal contamination.

2.5. Conclusions

From drilodefensins to betaine analogs, earthworms are equipped with an abundance of unique metabolites. Still, unassigned resonances remain in earthworm NMR spectra, particularly in the aromatic region where conjugation and proton position on the aromatic ring results in complex *J*-coupling patterns with coupling patterns that are difficult to distinguish in complex mixtures. Separation techniques to target aromatic compounds can be employed to simplify spectra, but due to resonance overlap, other analytical techniques like mass spectrometry may be more useful in identifying aromatic unknowns. Performing measurements at higher magnetic fields improves both NMR spectral resolution and sensitivity and may be useful in the identification of minor species. Complexity due to resonance overlap can also be overcome through isotopic labeling of specific functional groups, for example, ^{13}C-formlyation of amines and ^{15}N-ethanolamine tagging of carboxylates have improved the sensitivity and resolution of metabolite assignments using 2D NMR experiments.[43-44] Less resonance overlap is observed in ^{13}C and ^{15}N NMR compared to ^{1}H NMR, and recent advances in dissolution dynamic nuclear polarization have improved ^{13}C sensitivity, thus increasing the feasibility of using 1D ^{13}C

experiments in metabolic profiling.[45] Other tactics used to identify metabolites include chemical separation or orthogonal analytical techniques like Fourier transform-infrared spectroscopy and mass spectrometry.

Mass spectrometry has not been widely applied in earthworm metabolomic research but could offer insights into additional biochemical pathways and new metabolites. Gas chromatography-mass spectrometry (GC-MS) is commonly used in untargeted metabolic profiling experiments.[46] GC-MS has been employed to probe metabolic perturbations by xenobiotics in whole-worm extracts, revealing insights into fatty acids, low-level amino acids, carbohydrates, and lipids not detected in ^{1}H NMR spectra,[31, 47-51] yet GC-MS analysis of CF and CC extracts has yet to be extensively reported. Chapter 4 explores the use of ^{1}H NMR and GC-MS to examine the metabolic impacts of chloroacetanilide herbicide exposure on the CF metabolic profile.

Liquid chromatography-mass spectrometry (LC-MS) has been used to aid metabolite elucidation, but surprisingly we were unable to identify published reports of targeted or untargeted LC-MS experiments used to probe the earthworm metabolome. Targeted LC-MS can be used to survey chemical classes or biochemical pathways, like phosphorylated sugars, aiding metabolite elucidation or providing insights into specific metabolic functions.[52] Chapter 5 demonstrates the utility of targeted LC-MS analysis along with ^{1}H NMR to identify potential biomarkers of chlorothalonil exposure. In addition, extending the use of LC-MS in earthworm metabolomics studies can help elucidate unassigned resonances in the ^{1}H NMR spectra and give new insights into the biochemical function of unique earthworm metabolites.

As demonstrated in this study, NMR offers the means to quantify a diverse range of metabolites in earthworms for systems toxicology research or environmental monitoring. Our work, and that of similar studies, provide a platform to enhance the understanding of earthworm metabolism and illustrate the potential for high-throughput NMR analysis to deliver a quick snapshot of earthworm, soil, and ecological health.

2.6. References

1. Griffith, C. M.; Baig, N.; Seiber, J. N., Contamination from industrial toxicants. In *Handbook of food chemistry*, Cheung, P. C. K.; Mehta, B. M., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; pp 719-751.

2. Griffith, C. M.; Woodrow, J. E.; Seiber, J. N., Environmental behavior and analysis of agricultural sulfur. *Pest Manag. Sci.* **2015,** *71* (11), 1486-1496.

3. Bundy, J. G.; Osborn, D.; Weeks, J. M.; Lindon, J. C.; Nicholson, J. K., An nmr-based metabonomic approach to the investigation of coelomic fluid biochemistry in earthworms under toxic stress. *FEBS Lett.* **2001,** *500* (1-2), 31-35.

4. Yuk, J.; Simpson, M. J.; Simpson, A. J., Coelomic fluid: A complimentary biological medium to assess sub-lethal endosulfan exposure using h-1 nmr-based earthworm metabolomics. *Ecotoxicology* **2012,** *21* (5), 1301-1313.

5. Yuk, J.; Simpson, M. J.; Simpson, A. J., 1-d and 2-d nmr-based metabolomics of earthworms exposed to endosulfan and endosulfan sulfate in soil. *Environ. Pollut.* **2013,** *175*, 35-44.

6. Bundy, J. G.; Spurgeon, D.; Svendsen, C.; Hankard, P.; Osborn, D.; Lindon, J.; Nicholson, J., Earthworm species of the genus eisenia can be phenotypically differentiated by metabolic profiling. *FEBS Lett.* **2002,** *521*, 115 - 120.

7. Brown, S. A. E.; Simpson, A. J.; Simpson, M. J., Evaluation of sample preparation methods for nuclear magnetic resonance metabolic profiling studies with eisenia fetida. *Environ. Toxicol. Chem.* **2008,** *27* (4), 828-836.

8. Bax, A.; Davis, D. G., Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985,** *65* (2), 355-360.

9. Liu, M.; Mao, X.-a.; Ye, C.; Huang, H.; Nicholson, J. K.; Lindon, J. C., Improved watergate pulse sequences for solvent suppression in nmr spectroscopy. *J. Magn. Reson.* **1998,** *132* (1), 125-129.

10. Davis, A. L.; Laue, E. D.; Keeler, J.; Moskau, D.; Lohman, J., Absorption-mode two-dimensional nmr spectra recorded using pulsed field gradients. *J. Magn. Reson.* **1991,** *94* (3), 637-644.

11. Piantini, U.; Sorensen, O. W.; Ernst, R. R., Multiple quantum filters for elucidating nmr coupling networks. *J. Am. Chem. Soc.* **1982,** *104* (24), 6800-6801.

12. Aue, W. P.; Karhan, J.; Ernst, R. R., Homonuclear broad band decoupling and two-dimensional j-resolved nmr spectroscopy. *J. Chem. Phys.* **1976,** *64* (10), 4226-4227.

13. Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M., Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation nmr spectroscopy. *J. Magn. Reson.* **1991,** *93* (1), 151-170.

14. Kay, L.; Keifer, P.; Saarinen, T., Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992,** *114* (26), 10663-10665.

15. Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W., Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **1993,** *31* (3), 287-292.

16. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C., A general enhancement scheme in heteronuclear multidimensional nmr employing pulsed field gradients. *J. Biomol. NMR* **1994,** *4* (2), 301-306.

17. Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the madison metabolomics consortium database. *Nat Biotech* **2008,** *26* (2), 162-164.

18. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., Hmdb 3.0—the human metabolome database in 2013. *Nucleic Acids Research* **2013,** *41* (D1), D801-D807.

19. Liebeke, M.; Bundy, J. G., Biochemical diversity of betaines in earthworms. *Biochem. Bioph. Res. Co,* **2013,** *430* (4), 1306-1311.

20. Liebeke, M.; Strittmatter, N.; Fearn, S.; Morgan, A. J.; Kille, P.; Fuchser, J.; Wallis, D.; Palchykov, V.; Robertson, J.; Lahive, E.; Spurgeon, D. J.; McPhail, D.; Takats, Z.; Bundy, J. G., Unique metabolites protect earthworms against plant polyphenols. *Nat Commun* **2015,** *6*, 7869.

21. Xiong, L.; Ping, L.; Yuan, B.; Wang, Y., Methyl group migration during the fragmentation of singly charged ions of trimethyllysine-containing peptides: Precaution of using ms/ms of singly charged ions for interrogating peptide methylation. *J Am Soc Mass Spectrom* **2009,** *20* (6), 1172-1181.

22. Naresh Chary, V.; Dinesh Kumar, C.; Vairamani, M.; Prabhakar, S., Characterization of amino acid-derived betaines by electrospray ionization tandem mass spectrometry. *J. Mass. Spectrom.* **2012,** *47* (1), 79-88.

23. Patti, A.; Morrone, R.; Chillemi, R.; Piattelli, M.; Sciuto, S., Biosynthetic relationships between sulfonium and n-methylated compounds in the red alga vidalia volubilis. *J. Nat. Prod.* **1992,** *55* (1), 53-57.

24. Rochfort, S.; Wyatt, M. A.; Liebeke, M.; Southam, A. D.; Viant, M. R.; Bundy, J. G., Aromatic metabolites from the coelomic fluid of eisenia earthworm species. *Eur. J. Soil. Bio.* **2017,** *78*, 17-19.

25. Maraldo, K.; Ravn, H.; Slotsbo, S.; Holmstrup, M., Responses to acute and chronic desiccation stress in enchytraeus (oligochaeta: Enchytraeidae). *J. Comp. Physiol. B* **2009,** *179* (2), 113-123.

26. Bayley, M.; Overgaard, J.; Hoj, A. S.; Malmendal, A.; Nielsen, N. C.; Holmstrup, M.; Wang, T., Metabolic changes during estivation in the common earthworm aporrectodea caliginosa. *Physiol. Biochem. Zool.* **2010,** *83* (3), 541-550.

27. Al Temimi, A. H. K.; Pieters, B. J. G. E.; Reddy, Y. V.; White, P. B.; Mecinovic, J., Substrate scope for trimethyllysine hydroxylase catalysis. *Chem. Commun.* **2016,** *52* (87), 12849-12852.

28. Vaz, F. M.; Wanders, R. J. A., Carnitine biosynthesis in mammals. *Biochem. J.* **2002,** *361* (Pt 3), 417-429.

29. Polticelli, F.; Salvi, D.; Mariottini, P.; Amendola, R.; Cervelli, M., Molecular evolution of the polyamine oxidase gene family in metazoa. *BMC Evol. Biol.* **2012,** *12* (1), 1-14.

30. Hamana, K.; Hamana, H.; Shinozawa, T., Alterations in polyamine levels of nematode, earthworm, leech and planarian during regeneration, temperature and osmotic stresses. *Comp. Biochem. Phys. B* **1995,** *111* (1), 91-97.

31. Baylay, A. J.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L.; Swain, S.; Sturzenbaum, S.; Jones, O. A. H., A metabolomics based test of independent action and concentration addition using the earthworm lumbricus rubellus. *Ecotoxicology* **2012,** *21* (5), 1436-1447.

32. Givaudan, N.; Wiegand, C.; Le Bot, B.; Renault, D.; Pallois, F.; Llopis, S.; Binet, F., Acclimation of earthworms to chemicals in anthropogenic landscapes, physiological mechanisms and soil ecological implications. *Soil Biol. Biochem.* **2014,** *73*, 49-58.

33. Lankadurai, B. P.; Furdui, V. I.; Reiner, E. J.; Simpson, A. J.; Simpson, M. J., 1h nmr-based metabolomic analysis of sub-lethal perfluorooctane sulfonate exposure to the earthworm, eisenia fetida, in soil. *Metabolites* **2013,** *3* (3), 718-40.

34. Ji, C. L.; Wu, H. F.; Wei, L.; Zhao, J. M.; Lu, H. J.; Yu, J. B., Proteomic and metabolomic analysis of earthworm eisenia fetida exposed to different concentrations of 2,2 ',4,4 '-tetrabromodiphenyl ether. *J. Proteomics* **2013,** *91*, 405-416.

35. Kobayashi, H.; Ohta, N.; Umeda, M., Biology of lysenin, a protein in the coelomic fluid of the earthworm eisenia foetida. In *International review of cytology*, Academic Press: 2004; Vol. Volume 236, pp 45-99.

36. Hanušová, R.; Bilej, M.; Brys, L.; De-Baetselier, P.; Beschin, A., Identification of a coelomic mitogenic factor in eisenia foetida earthworm. *Immunology Letters* **1999,** *65* (3), 203-211.

37. Liu, Y.-Q.; Sun, Z.-J.; Wang, C.; Li, S.-J.; Liu, Y.-Z., Purification of a novel antibacterial short peptide in earthworm eisenia foetida. *Acta Biochimica et Biophysica Sinica* **2004,** *36* (4), 297-302.

38. Saint-Denis, M.; Labrot, F.; Narbonne, J. F.; Ribera, D., Glutathione, glutathione-related enzymes, and catalase activities in the earthworm eisenia fetida andrei. *Arch. Environ. Contam. Toxicol.* **1998,** *35* (4), 602-614.

39. Bilej, M.; Procházková, P.; Šilerová, M.; Josková, R., Earthworm immunity. In *Invertebrate immunity*, Söderhäll, K., Ed. Springer US: 2010; Vol. 708, pp 66-79.

40. Mazur, A. I.; Klimek, M.; Morgan, A. J.; Plytycz, B., Riboflavin storage in earthworm chloragocytes and chloragocyte-derived eleocytes and its putative role as chemoattractant for immunocompetent cells. *Pedobiologia* **2011,** *54, Supplement* (0), S37-S42.

41. Santocki, M.; Falniowski, A.; Plytycz, B., Restoration of experimentally depleted coelomocytes in juvenile and adult composting earthworms eisenia andrei e. Fetida and dendrobaena veneta. *Appl. Soil Ecol.* **2016,** *104*, 163-173.

42. Plytycz, B.; Lis-Molenda, U.; Cygal, M.; Kielbasa, E.; Grebosz, A.; Duchnowski, M.; Andre, J.; Morgan, A. J., Riboflavin content of coelomocytes in earthworm (dendrodrilus rubidus) field populations as a molecular biomarker of soil metal pollution. *Environ. Pollut.* **2009,** *157* (11), 3042-3050.

43. Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S.; Raftery, D., Chemoselective 15n tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Anal. Chem.* **2009,** *81* (12), 4882-4888.

44. Ye, T.; Zhang, S.; Mo, H.; Tayyari, F.; Gowda, G. A. N.; Raftery, D., 13c-formylation for improved nuclear magnetic resonance profiling of amino metabolites in biofluids. *Anal. Chem.* **2010,** *82* (6), 2303-2309.

45. Bornet, A.; Maucourt, M.; Deborde, C.; Jacob, D.; Milani, J.; Vuichoud, B.; Ji, X.; Dumez, J.-N.; Moing, A.; Bodenhausen, G.; Jannin, S.; Giraudeau, P., Highly repeatable dissolution dynamic nuclear polarization for heteronuclear nmr metabolomics. *Anal. Chem.* **2016,** *88* (12), 6179-6183.

46. Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K., Comparison of gc-ms and nmr for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12* (2), 898-909.

47. Jones, O. A. H.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L., A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm lumbricus rubellus. *Chemosphere* **2008,** *71* (3), 601-609.

48. McKelvie, J. R.; Yuk, J.; Xu, Y. P.; Simpson, A. J.; Simpson, M. J., H-1 nmr and gc/ms metabolomics of earthworm responses to sub-lethal ddt and endosulfan exposure. *Metabolomics* **2009,** *5* (1), 84-94.

49. Mudiam, M. K. R.; Ch, R.; Saxena, P. N., Gas chromatography-mass spectrometry based metabolomic approach for optimization and toxicity evaluation of earthworm sub-lethal responses to carbofuran. *PLoS One* **2013,** *8* (12), 13.

50. Gillis, J. D.; Price, G. W.; Prasher, S., Lethal and sub-lethal effects of triclosan toxicity to the earthworm eisenia fetida assessed through gc-ms metabolomics. *J. Hazard. Mater.* **2017,** *323*, 203-211.

51. Ch, R.; Singh, A. K.; Pandey, P.; Saxena, P. N.; Reddy Mudiam, M. K., Identifying the metabolic perturbations in earthworm induced by cypermethrin using gas chromatography-mass spectrometry based metabolomics. *Sci. Rep.* **2015,** *5*, 15674.

52. Mathon, C.; Barding Jr, G. A.; Larive, C. K., Separation of ten phosphorylated mono-and disaccharides using hilic and ion-pairing interactions. *Anal. Chim. Acta* **2017,** *972*, 102-110.

CHAPTER THREE

Malylglutamate: Structural Elucidation and Exploration of Biological Function of a New Metabolite

Based in part on a paper published in the *Journal of Proteome Research*

J. Proteome Res. 2017, *16* (9), 3407-3417

Acknowledgement: I would like to thank Dr. Preston Williams for conducting the initial mass spectrometry experiments and data interpretation under the guidance of Dr. Yinsheng Wang. Additionally, I would like to thank Dr. David Martin and Abigale Fecue for conducting the malylglutamate synthesis and optical rotation measurements for this work. I would also like the thank Dr. Jay Kirkwood and the UCR Metabolomics Core Facility for use of their LC-MS for the chiral chromatography experiments, and analyzing the malylglutamate screening experiments on their LC-MS.

Abstract

A previously unknown set of resonances found at 7 μg/mg in the ^{1}H NMR spectra of earthworm extracts is identified herein as malylglutamate, a new metabolite. Two-dimensional NMR spectroscopy, high-resolution mass spectrometry, synthesis, chiral chromatography, and optical rotation was used to assign absolute stereochemistry as (–)-*β*-L-malyl-L-glutamate. With access to synthetic material, LC-MS/MS was used to detect malylglutamate in several invertebrates closely related on the phylogenetic tree. We postulate that malylglutamate acts as a glutamate/malate store, chelator, and anionic osmolyte and helps to provide electrolyte balance in earthworms and sought to explore its

biochemical function. Malylglutamate levels were unaffected by high salinity, dry soil, and cold stress, but decreased under high soil water content. This suggests that malylglutamate is unlikely to act as an osmolyte and cryoprotectant but may act as a store for malate/glutamate and provides charge balance. Malylglutamate was identified as a chelator of Ca^{2+}, Mn^{2+}, and Zn^{2+} using ^{1}H NMR, indicating that one of its main functions is to likely protect earthworms from metal burdens in soil.

3.1. Introduction

An intense set of metabolite resonances found in the ^{1}H NMR profiles of whole-earthworm, coelomic fluid, and coelomocyte extracts remained unassigned after initial NMR metabolic profiling described in Chapter 2. The unknown metabolite was found at approximately 7 μg/mg in whole-earthworms, and because of its high abundance, it was critical to identify the metabolite. This chapter follows the structural elucidation of a new metabolite, (–)-*β*-L-malyl-L-glutamate, using NMR spectroscopy, mass spectrometry, synthesis, chiral chromatography, and optical rotation. Following identification of malylglutamate, we sought to screen for its presence in invertebrates closely related on the phylogenetic tree, including other earthworm species, and probe its biological function.

3.2. Experimental Procedures

3.2.1. Malylglutamate Structure Elucidation

3.2.1.1. NMR Acquisition Parameters for Initial Assignment

NMR Acquisition Parameters for the initial malylglutamate structural elucidation are described in Sections 2.2.7 and 2.2.8 of Chapter 2. For the 90% H_2O samples, 60 μL of D_2O buffer was added to 400 μL of filtrate and 40 μL of ultrapure water prior to titration with deuterium chloride (Sigma Aldrich, St. Louis, MO, USA).

3.2.1.2. Direct Infusion – Tandem Mass Spectrometry Analysis

CF samples were loaded into a 500 μL syringe and introduced into an LTQ-Orbitrap Velos mass spectrometer equipped with a heated electrospray ionization source (Thermo Fisher Scientific, San Jose, CA, USA) at a flow rate of 10 μL per min.[1] The instrument was operated in the positive-ion mode and full-scan mass spectra in the range of *m/z* 50-1000 were acquired in the Orbitrap mass analyzer with the resolution of 100,000 at *m/z* 400. The precursors for the metabolites of interest were subjected to collision-induced dissociation (CID) to acquire the MS/MS spectra in the linear ion trap.

Data were analyzed manually using Xcalibur Qual Browser (Thermo Fisher Scientific, San Jose, CA, USA). Raw MS data were analyzed by searching the calculated exact mass of each metabolite of interest. Using the advanced processing within chromatogram ranges, data were smoothed using the boxcar algorithm with 7 points and the mass tolerance was user defined at 10 ppm. The plot type was set to mass range with the full MS scan filter applied using the MS detector and the ICIS peak algorithm was used.

The calculated exact mass of the metabolite of interest was placed in the range and the data were searched for analyte peaks. Each metabolite under investigation was searched individually using the same method. Signal was then averaged in the time range for the peaks found in the ion chromatogram to give MS for each metabolite.

3.2.1.3. Synthesis of Malylglutamate Isomers

Malylglutamate isomers were synthesized according to known literature procedures by Dr. David Martin and Abigale Fecue (UC Riverside).[2-5] Optical rotation was measured on a Rudolph Research Analytical Autopol IV Automatic Polarimeter. The synthetic methodology is described in Appendix A.

3.2.1.4. NMR Acquisition Parameters for Diastereomer Assignment

Prior to NMR analysis, samples were reconstituted in 200µL of 100 mM phosphate buffer (pD 7.45) in D_2O (D, 99.9%) containing 0.2 mM ethylenediaminetetraacetic acid-d_{16} (EDTA-d_{16}) and 0.25 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonic acid-d_6 (DSS-d_6) (Cambridge Isotope Laboratories, Tewksbury, MA) and transferred to a 3 mm NMR Tube (New Era Enterprises, Vineland, NJ). A Bruker Avance NMR spectrometer operating at 599.88 MHz and equipped with a SmartProbe was used to acquire one-dimensional and heteronuclear multiple bond correlation spectroscopy (HMBC) spectra.[6] ^1H survey spectra were acquired with 1D NOESY (noesypr1d) pulse sequence with presaturation during the 120 ms mixing time. Spectra were acquired at 25°C with 64 scans, 16 dummy scans, 2 s delay time, 1.95 s acquisition time, 13.9848 ppm spectral width, and 32768 points. The

FIDs were exported into MestReNova 12 (Santiago de Compostela, Spain), apodized by multiplication with an exponential function equivalent to 0.5 Hz line broadening, and zero-filled to 131072 points prior to Fourier transformation. The resulting spectra were manually phased and referenced to DSS-d_6 (0 ppm).

The HMBC (hmbcgplpndqf) experiment was conducted using an 80-worm pooled coelomocyte sample acquired at 25°C with 64 scans, 16 dummy scans, ^{1}H 90° pulse of 13.25 μs, ^{13}C 90° pulse of 13.00 μs, 65 ms evolution delay, proton-carbon coupling constant ($^{n}j_{CH}$) of 10 Hz, and a 2 s relaxation delay. Spectral widths of 13.9848 ppm and 210.0000 ppm were used with 2048 by 512 complex points acquired in F2 and F1, respectively. The HMBC spectrum was apodized using a sine function, baseline corrected with a third order polynomial in both dimensions, smoothed using a Savitzky-Golay filter with an order of 4 and width of 16 in both dimensions, zero-filled to 4096 by 4096 points, and manually phased.

3.2.1.5. Chiral Separation of L,L- and D,D-*β*-malylglutamate

A Waters UPLC coupled to a G2-XS Q-TOF mass spectrometer (Milford, MA, USA) was used for assignment of the correct diastereomer of *β*-malylglutamate in the natural sample. The separation was performed on an Asetc Chirobiotic® R 150 x 2.1 mm, 5 μm column (Sigma Aldrich, St. Louis, MO, USA) at 30°C using isocratic conditions of 98% ultrapure water containing 0.01% acetic acid (Fisher Scientific,) and 2% LiChrosolv® ethanol (EMD Millipore Corporation, Billerica, MA,USA) at a flow rate of 0.55 mL/min and the column maintained at 30°C. Electrospray ionization was used and β-

Malylglutamate was detected in positive ion mode at 264.071 *m/z* using a spray voltage of 35 V and source temperature of 150°C. MS/MS spectra were acquired for the *m/z* of interest.

L,L- and D,D-*β*-malylglutamate standards were dissolved in ultrapure water and diluted with methanol to 10 μM prior to a 1 μL injection. Coelomic fluid, coelomocyte, and whole-earthworm extracts were reconstituted in 1 mL 2:1.8 methanol:water, centrifuged for 1 min at 16100 *g* prior to analysis, and divided into 200 μL aliquots. In separate spiking experiments, 2 μL of 1.7 mM L,L-*β*-malylglutamate standard or 2 μL of 0.4 mM D,D-*β*-malylglutamate standard were added to aliquots of coelomic fluid prior to LC-MS/MS analysis. Chromatograms were processed using a mean smoothed function with a window size of ± 3 and 2 smooths. Mass spectra were deisotoped, lock mass corrected, and centered.

3.2.2. Exploration of Biological Function

3.2.2.1. Screening in Invertebrates using LC-MS

Live invertebrates were flash-frozen in liquid nitrogen upon arrival. Redworm (*Eisenia fetida*), nightcrawler (*Lumbricus terrestris*), blackworm (*Lumbriculus variegatus*), whiteworm (*Enchytraeus sp.*), leech (*Placobdella ornata*), black planaria (*Phagocata gracilis*), brown planaria (*Fugesia tigrine*), water flea (*Daphnia magna*), and water bear (*Hypsibius sp.*) were purchased from Ward's Science (Rochester, NY, US). *Caenorhabditis elegans* and mystery snails (*Pomacea bridgesi*) were purchased from Carolina Biological Supply (Burlington, NC, US). *C. elegans* arrived in agar and samples

were collected by setting up setting the agar in approximately 20 mL of ultrapure water for five minutes. The agar was removed, sample flash-frozen, and lyophilized. The entire jar of water bears and media were transferred to a 50 mL Falcon tube, flash-frozen, and lyophilized.

Lyophilized and weighed samples were homogenized by cryo-cooled bead-beating for three cycles of 10 s at 5 m/s (Omni International, Kennesaw, GA, US). At least three individuals of each species were pooled together prior to extraction. Dried homogenates were extracted with 2:1.8 methanol:ultrapure water at a ratio of 10 mg per 1 mL solvent. Samples were vortexed for 3 min, 4°C at 2000 rpm on a ThermoMixer® C (Eppendorf, Hauppauge, NY, US) and centrifuged at 16000 *g*, 4°C for 20 min. The supernatant was transferred to 200 µL LC glass inserts and stored at -80°C until analysis. Malylglutamate standards for the external calibration plot were prepared from a 1.7 mM stock in ultrapure water.

LC-MS was employed using a Waters I-class UPLC system coupled to a TQ-XS triple quadrupole mass spectrometer. Separations were conducted on a ZIC-pHILIC column (2.1 x 150 mm, 5 µm) (EMD Millipore) using (A) water with 15 mM ammonium bicarbonate adjusted to pH 9.6 with ammonium hydroxide and (B) acetonitrile as mobile phases. A 1 µL injection volume was used with a flow rate of 200 L/min and the column was held at 50° C. The gradient was as follows: 0 min, 90% B; 1.5 min, 90% B; 16 min, 20% B; 18 min, 20% B; 20 min, 90% B; 28 min, 90% B.

The MS was operated in selected reaction monitoring mode. Source and desolvation temperatures were held at 150° C and 500° C, respectively. Nitrogen was used

for the desolvation gas at 1000 L/h and cone gas at 150 L/h. Argon was used for the collision gas and was set to 0.15 mL/min. Capillary voltage was 1 kV in positive ion mode and 2 kV in negative ion mode. System stability was monitored by analyzing a quality control samples throughout the sample set, prepared by pooling together 6 μL of all samples into a single glass insert. To prevent artifactual metabolite changes between groups, samples were analyzed in random order. Peak integration was conducted with the open source software Skyline (University of Washington) by the staff at the UC Riverside Metabolomics Core Facility.[7]

3.2.2.2. Cold, Salinity and Water Stress

Earthworms were exposed to salinity, temperature, and water stress for 5-days in 1 L beakers containing 100 g of Magic Worm® bedding (Magic Products Inc., Amherst Junction, WI, USA). Soil was wetted with dechlorinated water for 24 h prior to adding worms, at a ratio of 2:1 water:soil, as suggested by the manufacture, for all experiments except the dry soil stress (1:1) and wet soil stress (3:1). All experiments were kept on the benchtop, except the cold stress group which was placed in the fridge (4°C) during the experiment. Salinity stress experiments were prepared by introducing 10, 40, or 160 mM NaCl in the water used to wet the soil. The control condition for this experiment is considered the 2:1 water:soil, held at room temperature, and 0 mM NaCl.

Earthworms were removed from lab cultures, rinsed, and patted dry prior to adding ten worms to each beaker. After 5-day exposure, earthworms were removed from soil, rinse, dried, flash-frozen in liquid nitrogen, lyophilized, and homogenized by cryo-cooled

bead-beating, as described in Section 3.2.2.1. Earthworm homogenates were extracted using 2:2:1.8 Chloroform (Macron Fine Chemical, Center Valley, PA, USA):Fischer Scientific Optima Methanol (Fair Lawn, NJ, USA):Ultrapure Water.[8] Aqueous solvents were first added to the tubes and vortexed at 2000 rpm, 4°C for 2 min on a ThermoMixer® C (Eppendorf, Hauppauge, NY, USA). A 600 μL aliquot of chloroform was added, and the samples were vortexed at 2000 rpm, 4°C for 2 min. The samples were centrifuged at 16000 *g*, 4°C for 20 min and left on ice for 10 min prior to transferring 800 μL to a fresh tube and dried overnight by speedvac.

Dried extracts were reconstituted in D_2O buffer and analyzed by NMR as described in Section 3.2.1.4. A Bruker 600 MHz NMR spectrometer using a 1D NOESY (noesypr1d) pulse sequence with presaturation during the 120 ms mixing time was used to acquire ^1H NMR spectra at 25°C with 128 scans, 16 dummy scans, 2 s delay time, 1.95 s acquisition time, 13.9848 ppm spectral width, and 32768 points. Spectra were processed as described in Section 3.2.1.4. A limited number of resonances were peak fit in MestReNova, which included acetate, alanine, betaine, fumarate, glutamate, glutamine, HEFS, *myo*-inositol, *scyllo*-inositol, α-ketoglutarate, malate, malylglutamate, trimethyllysine, and succinate. Peak fitting was conducted using a general Lorentzian peak shape, with a lower width constraint of 0.1 Hz, and upper constraint of 100 Hz, position constraint within ± 5%, maximum number of fine iterations of 100, and local minima filter of 5. The peak fitting results were exported into Excel and assembled into a single spreadsheet.

Samples were normalized by earthworm dry-weight prior to univariate analysis in IBM SPSS Statistics v24 (Armonk, NY, USA). Samples were subjected to the Sharpiro-

Wilks test of normality to assess skewness and Levene's test for homoscedasticity.[9] Analysis of variance (ANOVA) with Tukey's HSD *post hoc* test ($P = 0.05$) was used for pairwise comparisons of the salinity and water stress experiments. Two-sample t-tests ($P = 0.05$) assuming variance is equal between groups was used to compare control and cold stress experiments. Effect size using Glass' Δ was calculated for statistically significant metabolic changes.[10-11]

3.2.2.3. ^{1}H NMR Survey of Malylglutamate Chelation

CF samples were extruded as described in Chapter 2, Section 2.2.2 and pooled at a rate of 10 worms per mL of NaCl solution. The pooled CF was divided into 180 μL aliquots and 20 μL of stock or DSS-d_6 was added for a total volume of 200 μL, mixed thoroughly, and stored overnight in the fridge. Stock solutions of 100 μM $MnCl_2$ and 100 mM $CaCl_2$ and $ZnCl_2$ were prepared in D_2O (99.9%, D) containing 1 mM DSS-d_6. Mn^{2+} samples were titrated between 2 – 10 μM, while Ca^{+2} and Zn^{2+} samples were titrated between 1 – 10 mM. ^{1}H NMR survey spectra were acquired on a Bruker 600 MHz spectrometer using 1D NOESY (noesypr1d) pulse sequence with presaturation during the 120 ms mixing time. Spectra were acquired at 25°C with 512 scans, 16 dummy scans, 2 s delay time, 1.95 s acquisition time, 13.9848 ppm spectral width, and 32768 points. Spectra were processed as described in Section 3.2.1.4.

3.3. Results and Discussion

3.3.1. Malylglutamate Structure Elucidation

Malylglutamate, a previously unreported metabolite was elucidated in these studies using NMR and high-resolution mass spectrometry (HRMS). Two sets of unassigned resonances (Figure 3.1), denoted as 1H (green) and 2H (blue), were consistently observed in all *E. fetida* matrices. The resonance at 2.23 ppm (2Hγ; Figure 3.2A) was initially noted for its uncanny similarity in terms of coupling and intensity to that produced by the Hγ of the free amino acid glutamate (Figure 3.2A), suggesting the possibility of a glutamate (Glu) analog. The Glu-like 2H spin system was revealed in the TOCSY spectrum (Figure 3.2B), further supporting the hypothesis. Through comparison of numerous CF and CC spectra, we noticed a correlation in the intensity of the 2H and 1H resonances (Figure 3.2), which we initially hypothesized could arise from an aspartyl moiety as a dipeptide due to its shared ABX coupling pattern.

To investigate the presence of a dipeptide, we measured the TOCSY spectrum of a concentrated CF sample in 90% H_2O at pD 3.18, revealing a strong correlation between an amide proton and a spin system resembling glutamate (Figure 3.2C) further supporting the assignment to a dipeptide. Several dipeptides were obtained and their ^{1}H NMR spectra recorded under conditions similar to those of the CF and CC extracts. Glutamylaspartate (Glu-Asp) produced similar chemical shifts for the aspartyl moiety, but notably different chemical shift and *J*-coupling pattern for the GluHγ resonance, which collapsed to a triplet (Figure 3.2A) and is consistently observed in aspartylglutamate (Asp-Glu) and *N*-acetylaspartylglutamate (NAAG). The GluHγ of Asp-Glu and NAAG is shifted upfield to

Glu (Figure 2.6A) like the unassigned 2Hγ. In contrast, the GluHγ of Glu-Asp is shifted downfield. Based on this result, we surmised that the glutamate group must be present at the dipeptide C-terminus, and indeed the Asp-Glu and NAAG coupling patterns, besides GluHγ, produced similar resonances to those observed in our CF and CC spectra (Figure 3.2A). However, due to the lack of correspondence of the Asp chemical shifts of these dipeptides with the unassigned 1HABX resonances (Figure 3.2A), we were forced to consider other alternatives.

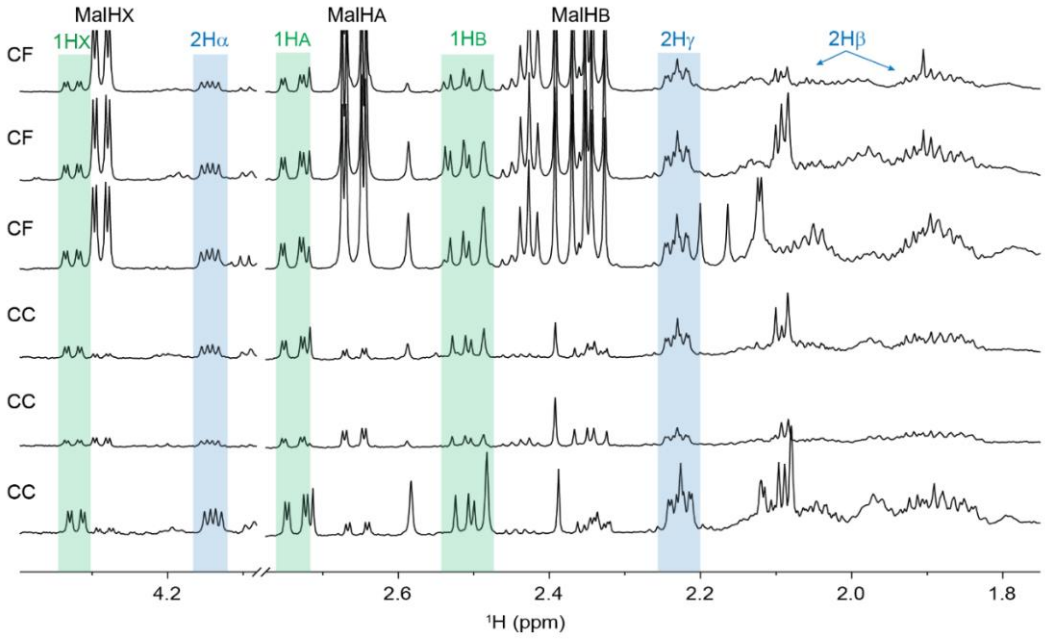


Figure 3.1. ^{1}H NMR spectra of individual worm coelomic fluid (CF) and coelomocyte (CC) samples with malate and the unassigned resonances 1H (green) and 2H (blue) labeled. These spectra illustrate the correlation in the intensity of the glutamate-like 2H (blue) resonances and malate-like 1H (green) resonances. The similarity in chemical shift of malate and 1H (green) is also illustrated.

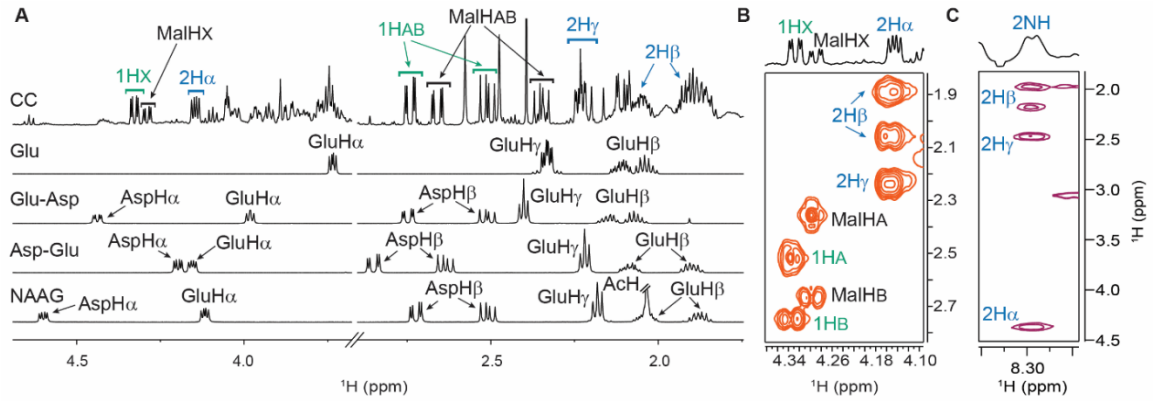


Figure 3.2. NMR spectra of unassigned resonances and comparison of closely related structures. A) ^{1}H NMR spectra of CC and the authentic standards of glutamate (Glu); glutamylaspartate (Glu-Asp); aspartylglutamate (Asp-Glu); and *N*-acetylaspartylglutamate (NAAG) are stacked below and resonance position is annotated. The two sets of unassigned resonances are annotated in the CC spectrum as 1H (green) and 2H (blue), and the malate (Mal) resonances are annotated in red. B) TOCSY spectrum of CC annotated with the correlated spin systems of 1H, 2H, and Mal. C) TOCSY spectrum of pooled CF sample at pD 3.18, in 90% H_2O revealing a strong correlation between an amide proton and a spin system resembling glutamate.

We proceeded to consider the similarity of the unassigned ABX resonances with malate, as annotated in Figure 3.2AB, suggesting that the dipeptide could be malylglutamate (Mal-Glu). Supporting this hypothesis, the ^{1}H and ^{13}C chemical shifts and 2Hγ *J*-coupling of our unknown closely resembled that of *β*-citrylglutamate, described in Collard, et al. [12] We were unable to identify reference to Mal-Glu in the literature or to

locate a commercial source of the compound for verification by NMR. Therefore, we turned to HRMS to explore our hypothesis that the unknown metabolite could be Mal-Glu.

A diluted CF sample was directly infused into the mass spectrometer (DI-MS), yielding the Mal-Glu parent ion at *m/z* 264.0667 (the calculated *m/z* 264.0719 for the $[M + H]^+$ ion; Figure 3.3A). Ions in the MS/MS (Figure 3.3B) of the $[M + H]^+$ ion were assigned: *m/z* 246.1 $[M - H_2O]^+$; 227.9 $[M - 2H_2O]^+$; 220.1 $[M - CO_2]^+$; 190.0 $[M – CH_3CH_3COOH]^+$; and 146.0 $[M – CHOCH(OH)CH_2COOH]^+$. The fragments at *m/z* 146.0, 190.0, and 220.1 show the loss of three distinct groups around the single central nitrogen, confirming the presence of the amide bond and C-terminus position of the Glu residue. The peak at *m/z* 232.1 is thought to be arise from a co-fragmenting species with the same *m/z* value as the newly discovered analyte due to the nature of the complex matrix of CF. Taken together, the NMR and HRMS results provide a high level of confidence that the unknown metabolite is Mal-Glu, first reported herein.

3.3.1.1. Synthesis and ^{1}H NMR Comparisons

Following up on the identification of malylglutamate using NMR and DI-MS/MS, we sought to confirm the proposed structure of malylglutamate (**1**, Figure 3.4) and fully assign the relative and absolute stereochemistry.[13] Due to complexity of ^{1}H NMR spectra of earthworm extracts and the difficulty in isolating a pure sample for further studies, we turned our attention to synthesis using a traditional peptide coupling approach (Appendix A). This strategy allows the incorporation of either L- or D-isomers of both glutamic and malic acids to determine both relative and absolute configuration.

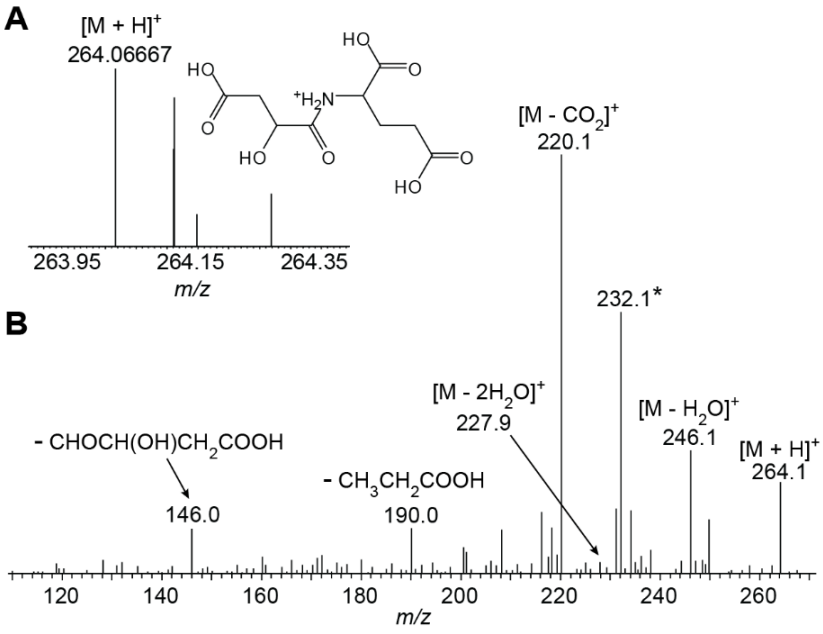


Figure 3.3. Confirmation of malylglutamate in a diluted CF sample using positive-ion ESI-MS and MS/MS. A) Positive-ion ESI-MS of malylglutamate (the calculated *m/z* 264.0719 for $[M + H]^+$ ion) with its structure and B) annotated MS/MS for the $[M + H]^+$ ion of malylglutamate. The * indicates a peak from an ion co-fragmenting with $[M + H]^+$ ion.

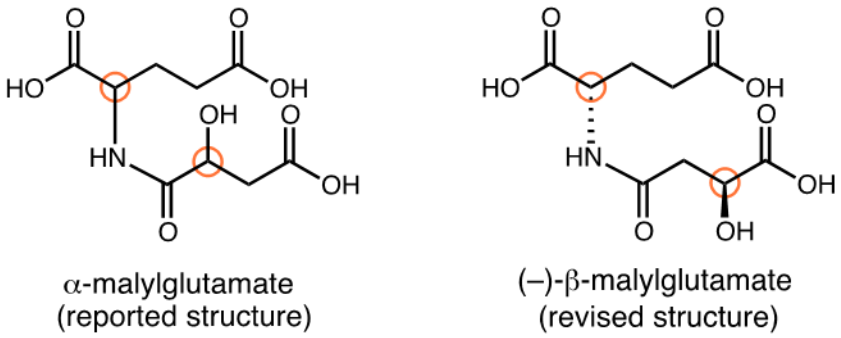


Figure 3.4. Structures of malylglutamate isomers and structurally similar metabolites: *α*-malylglutamate (initially reported structure)[13] and (–)-*β*-L-malyl-L-glutamate (revised structure).

Comparison of the ^{1}H NMR spectra of the synthetic L,L- and L,D-α-malylglutamate in D_2O with a naturally derived sample of malylglutamate showed that the compounds were not identical (Figure 3.5). In particular, the resonances of proton 2Hα were slightly shifted downfield in both synthetic molecules and proton 1Hx was significantly shifted downfield in both isomers. To ascertain whether the inconsistencies between the synthetic standards and native sample were due to matrix or other effects, heteronuclear multiple bond correlation spectroscopy (HMBC) was employed to explore long-range correlations between protons and carbons.[6] A pooled earthworm coelomocyte sample was used to conduct the analyses (Figure 3.6). Correlations between both of glutamate's carbonyls with the Hαβγ protons nicely show the correct assignment. Looking at the malyl moiety, equal intensities were observed between the carbonyls and 1Hx proton which did not aid with the assignment, labeled as 1Hx/CONH and 1Hx/COOH peaks in Figure 3.6; however, a very intense correlation between the amide carbonyl and 1HAB protons was observed, noted as 1HAB/CONH (Figure 3.6). This suggests that 1HAB protons were positioned directly next to the amide carbonyl and indicates that the malyl moiety was attached to glutamate in the β position (i.e., β-malylglutamate, Figure 3.4).

3.3.1.2. Reassignment and Elucidation of Absolute Configuration

Isomers of β-malylglutamate were synthesized and their ^{1}H NMR spectra were compared in D_2O buffer to a CC extract sample. While the D,L-β-malylglutamate isomer differed in *J*-coupling of the Hγ protons (Figure 3.7), the L,L-β-malylglutamate isomer provided an excellent match in chemical shift and *J*-coupling of the resonances in the

earthworm extract. To finalize the stereochemical assignment of *β*-malylglutamate, we sought to determine the absolute configuration. D,D-*β*-malylglutamate was synthesized using identical methods as described in Appendix A. A chiral chromatography method was developed to separate L,L- and D,D-*β*-malylglutamate using LC-MS. Excellent separation was achieved, yielding retention times of 1.61 min for L,L- and 8.28 min for D,D-*β*-malylglutamate and identical MS/MS spectra (Figure 3.8). The D,D-*β*-malylglutamate standard contained a second peak at 3.14 min with identical *m/z* as our standards; however, the MS/MS spectrum of the unknown peak was inconsistent with L,L- and D,D-*β*-malylglutamate standards (Figure 3.8), indicating that it is an isomer of 264.071 *m/z*. To identify the enantiomer of malylglutamate present in the natural sample, coelomic fluid, coelomocyte, and whole-earthworm extracts were injected on the LC-MS and the retention times and MS/MS spectra revealed L,L-*β*-malylglutamate as the correct diastereomer (Figure 3.8A). To confirm the assignment, L,L- and D,D-*β*-malylglutamate were spiked into separate coelomic fluid samples where an increase in the L,L-*β*-malylglutamate peak intensity was observed in the L,L-*β*-malylglutamate spiked sample (Figure 3.8B). A new peak was observed in the D,D-*β*-malylglutamate spiked sample and the L,L-*β*-malylglutamate peak intensity stayed consistent with the original sample (Figure 3.8C). The specific rotation of synthetic L,L-*β*-malylglutamate was determined to be $[\alpha]_D^{25}$ –13.6 (*c* 1.0, CH_3OH), thereby establishing the natural product to be (–)-*β*-L-malyl-L-glutamate.

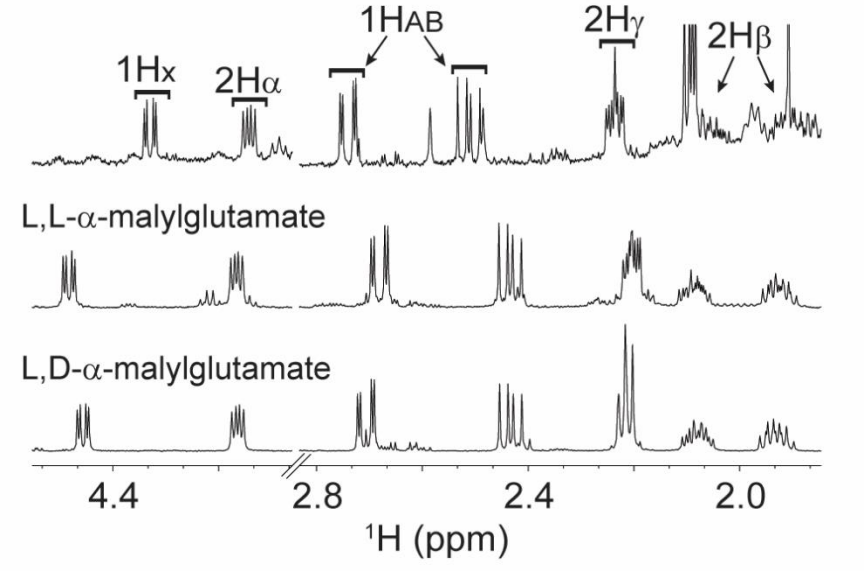


Figure 3.5. NMR spectrum an earthworm coelomocyte sample (top) compared to L,L- and L,D-α-malylglutamate (Appendix A).

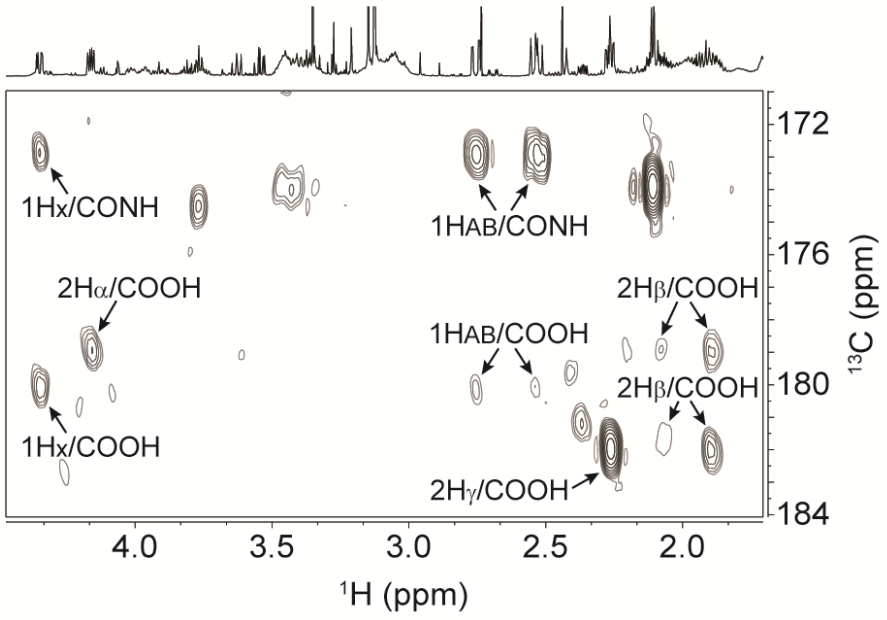


Figure 3.6. HMBC spectrum of 80-worm pooled coelomocyte sample. The peaks of malylglutamate are annotated and the 1HAB/CONH peaks suggest a reassignment of the structure to β-malylglutamate.

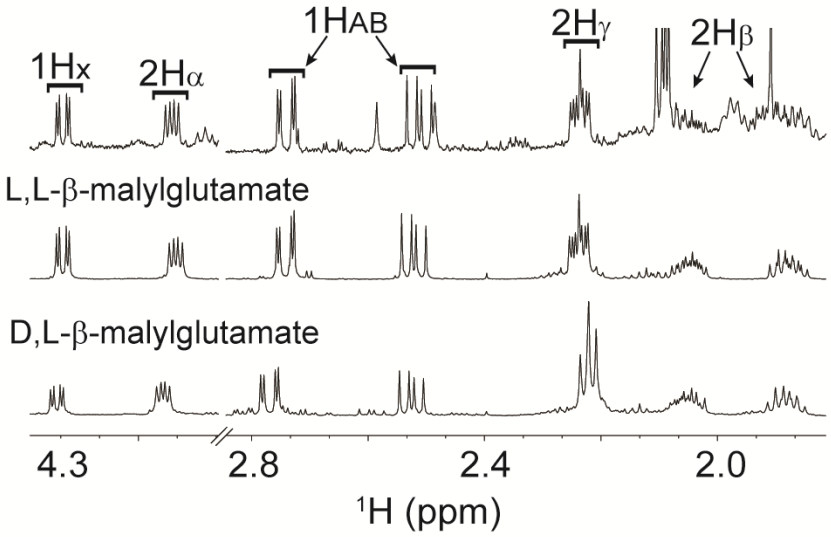


Figure 3.7. NMR spectrum an earthworm coelomocyte sample (top) compared to L,L- and D,L-*β*-malylglutamate (Appendix A).

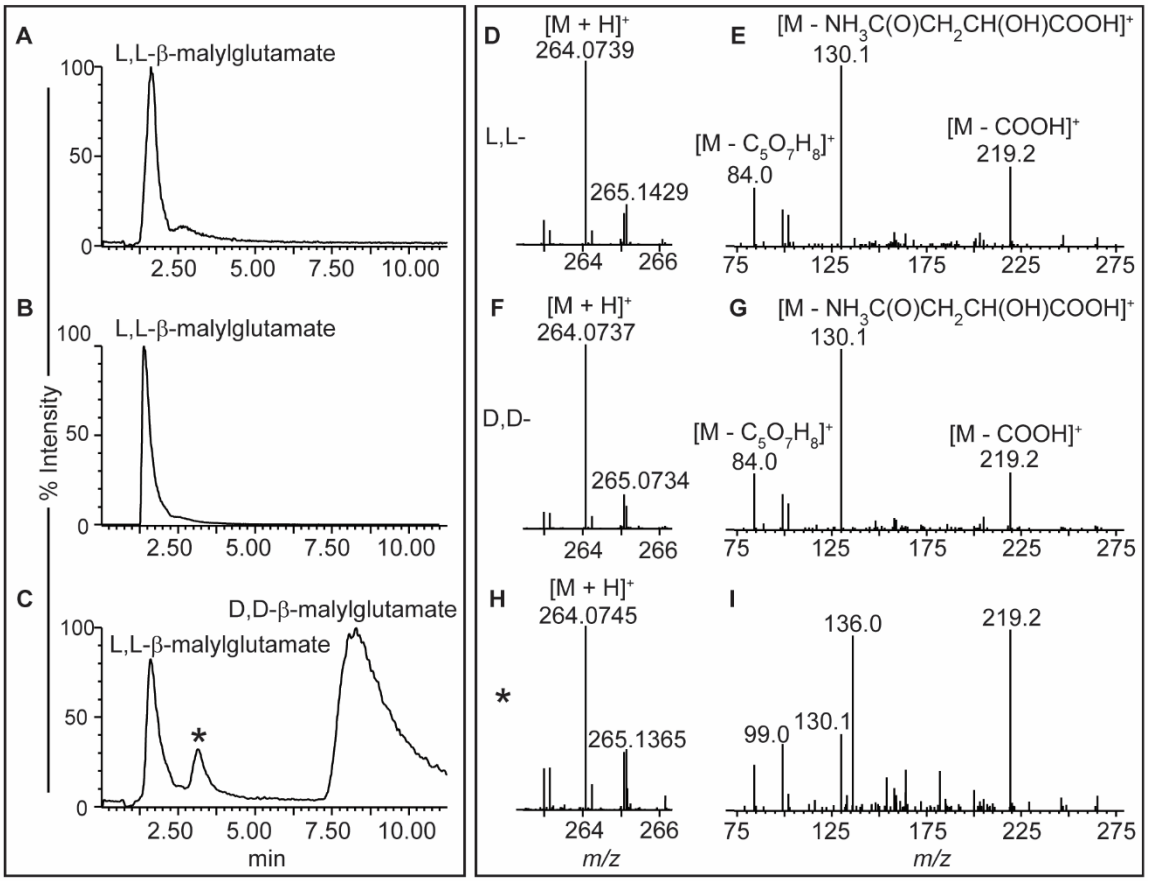


Figure 3.8. The chiral separation of L,L- and D,D-*β*-malylglutamate isomers using LC-MS at *m/z* 264.071. Chromatograms of (A) CF sample; (B) CF sample spiked with L,L-*β*-malylglutamate; and (C) CF sample spiked with D,D-*β*-malylglutamate, where *denotes an unknown isomer of $[M + H]^+$ ion. Positive-ion ESI-MS and annotated MS/MS of malylglutamate (calculated *m/z* 264.0719 for $[M + H]^+$ ion) are shown for each isomer: (D) MS^1 of L,L-*β*-malylglutamate isomer and (E) its annotated MS/MS spectrum for the $[M + H]^+$ ion; (F) MS^1 of D,D-*β*-malylglutamate isomer and (G) its annotated MS/MS spectrum for the $[M + H]^+$ ion; and (H) MS^1 of the unknown isomer and (I) its annotated MS/MS spectrum for the $[M + H]^+$ ion.

3.3.2. Exploration of Biological Function

Malylglutamate is structurally analogous to *β*-citrylglutamate and NAAG (Figure 3.9), and we hypothesize its synthesis and hydrolysis occurs via parallel pathways (Figure 3.10). NAAG and *β*-citrylglutamate are synthesized by ligases homologous to *Escherichia coli* RIMKLA and RIMKLB, respectively.[12] Whereas citrate is not a suitable substrate for RIMKLA, RIMKLB can ligate *N*-acetylaspartate and glutamate at similar rates to citrate and glutamate. The authors were not surprised by this finding since citrate and *N*-acetylaspartate are nearly isosteric. This suggest RIMKLB is less specific than RIMKLA, and it is feasible that the structurally-similar malate could be a suitable substrate in an environment or system with higher malate levels compared to citrate, as is the case with *E. fetida* CF. *β*-Citrylglutamate is cleaved into citrate and glutamate by *β*-citrylglutamate hydrolase (i.e., glutamate carboxypeptidase 3), a glycosylated, membrane-bound ecotoenzyme that is active only extracellularly (Figure 3.10).[14] Malate levels are strikingly higher in the CF compared to the CC (Figure 3.1), suggesting that malylglutamate hydrolysis may occur by this or a related ectoenzyme in the coelomic cavity, presumably to feed the TCA cycle and provide energy to muscle and CC.

Malylglutamate plausibly serves as a TCA cycle store that can be hydrolyzed to malate and glutamate when energy demand increases. Malylglutamate is also homologous to the anionic osmolyte 1,3,4,6-tetracarboxyhexane (Figure 3.9D), found in archaea, that helps counter K^+ concentrations under osmotic stress.[15] Malylglutamate is found in an approximately 1:1 ratio with TMO and TML [i.e., Mal-Glu:(TMO+TML)], suggesting that malylglutamate assists homeostasis by maintaining charge balance with the cationic

betaine analogs, in addition to acting as an anionic osmolyte to aid varying soil moisture contents. Lastly, we hypothesize that malylglutamate is a chelator, like *β*-citrylglutamate which can sequester iron and copper (Figure 3.9A).[16-17] *β*-Citrylglutamate may act as a chaperone for iron to aconitase and as a protective complex against NO-induced aconitase inhibition.[18-19] Malylglutamate may form similar complexes to aid biochemical processes and/or protect the worm against metal burdens. The physiological functions of malylglutamate were explored further to understand its presence in invertebrates closely related on the phylogenetic tree (Section 3.3.3); investigate its potential role as an osmolyte and malate/glutamate store (Section 3.3.4); and explore its ability to chelate metals (Section 3.3.5).

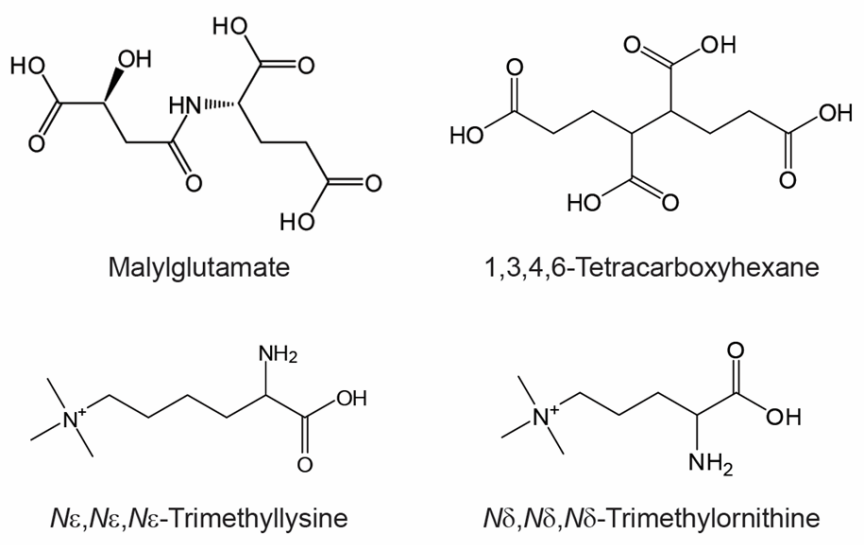


Figure 3.9. Chemical structures of a) *β*-malylglutamate; b) *β*-citrylglutamate; c) *N*-acetylaspartylglutamate (NAAG); and d) 1,3,4,6-tetracarboxyhexane.

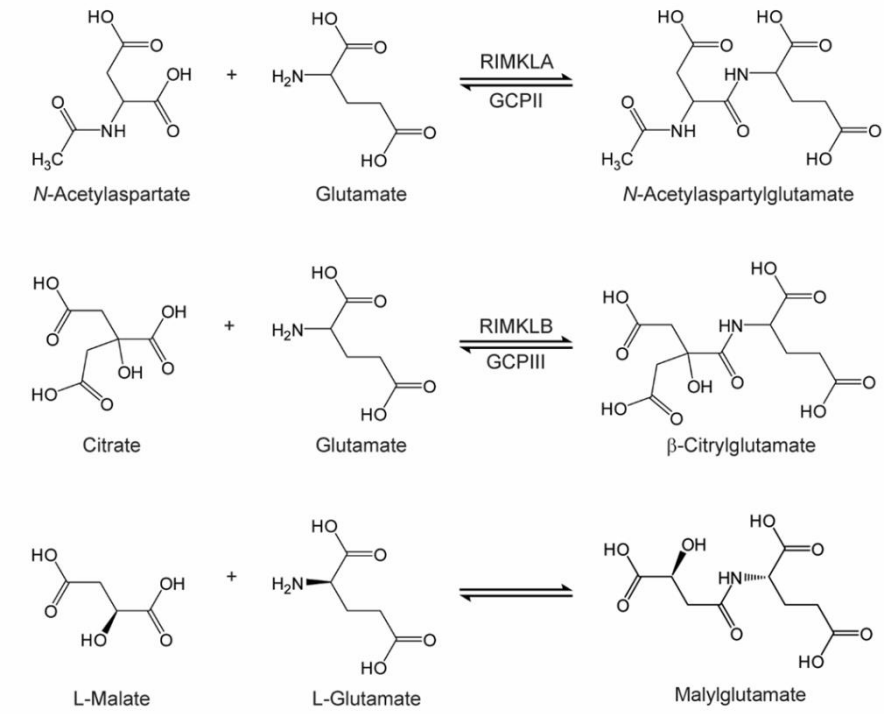


Figure 3.10. Biosynthetic pathways of *N*-acetylaspartylglutmate and *β*-citrylglutamate and suspected biosynthetic pathway of malylglutamate.

3.3.3. Screening for Malylglutamate in Invertebrates

A major question that arose from identifying an abundant, new metabolite was whether it was present in other species. Synthesis and elucidation of the absolute configuration of malylglutamate allowed us to employ LC-MS/MS to screen for malylglutamate in other organisms. The phylogenetic tree (Figure 3.11) was used as tool to select species closely related to *E. fetida.* Several members of the Annelida phylum, where earthworms reside, were selected: nightcrawler (*Lumbricus terrestris*), blackworm (*Lumbriculus variegatus*), whiteworm (*Enchytraeus sp.*), leech (*Placobdella ornata*). Mystery snail (*Pomacea bridgesi*) and water flea (*Daphnia magna*) are members of the

Mollusca and Arthropoda phylums, respectively. The Platyhelminthes black planaria (*Phagocata gracilis*) and brown planaria (*Fugesia tigrine*), the Nematoda *Caenorhabditis elegans*, and Tardigrada water bear (*Hypsibius sp.*) were also selected. An LC-MS/MS method (Figure 3.11) was optimized and malylglutamate was detected at a retention time of 13.8 min. In addition to the parent ion of 264.071 *m/z*, the MS/MS transitions of *m/z* 84, 102, 116, and 130 were monitored for further confirmation. An external calibration curve between 0 – 3.33 μM was used to calculate relative concentrations of malylglutamate in the invertebrate extracts. As summarized in Table 3.1, malylglutamate was detected at several species between 0.99 and >876.46 ng/mg. The peak intensities of malylglutamate in redworms were above the linear range (i.e., >876.46 ng/mg) of the calibration curve. Malylglutamate was detected in nightcrawlers and blackworms at varying concentration, while it was not detected in white worms, highlighting that its abundance varies in the Annelida phylum. Malylglutamate was not detected in Mollusca, but was detected in black planaria, *C. elegans*, and water bears. Malylglutamate was potentially detected in brown planaria and leeches, but the assignment is uncertain due to interference by co-eluting metabolites. This work suggests that malylglutamate is not limited to the Annelida phylum, and further screening should be conducted to evaluate its ubiquity in nature.

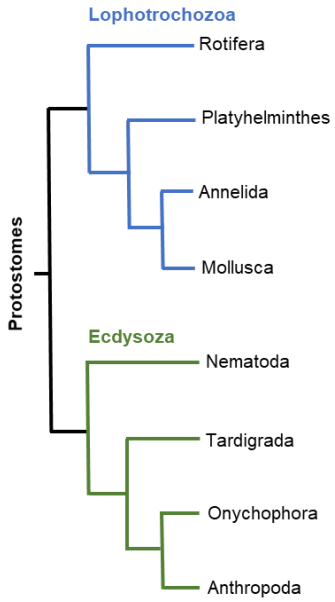


Figure 3.11. The phylogenetic tree of Protostomes.

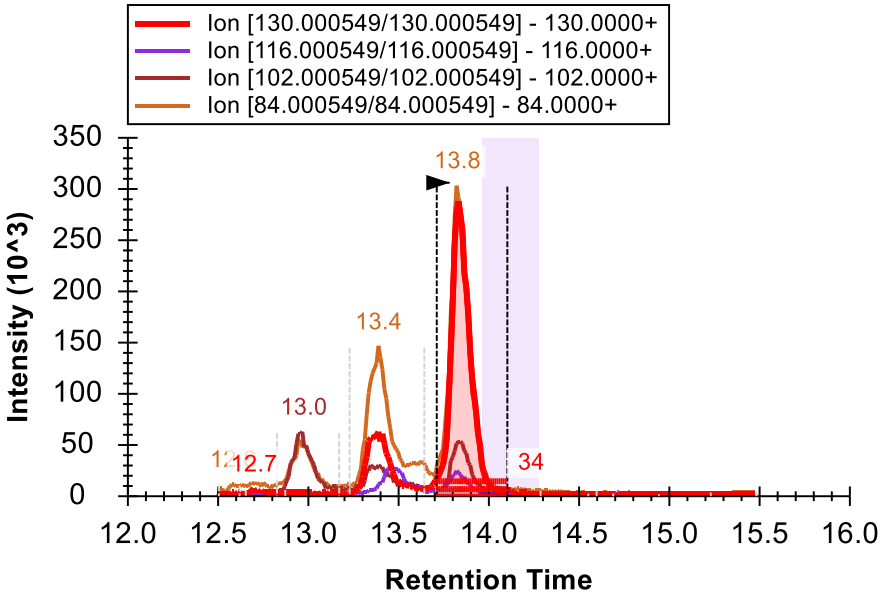


Figure 3.12. LC-MS/MS chromatogram of malylglutamate at a retention time of 13.8 min in a *C. elegans* extract. The MS/MS transitions that were used for confirmation are shown.

Table 3.1. Summary of malylglutamate detection and relative concentration per dry-weight of sample (mg) is reported using LC-MS/MS. NQ not quantifiable; ND = not detected.

Species	Phylum	[ng/mg]
Redworm (*Eisenia fetida)*	Annelid	>876.46
Nightcrawler (*Lumbricus terrestris*)	Annelid	222.44
Blackworm (*Lumbriculus variegatus*)	Annelid	16.63
White worm (*Enchytraeus sp*.)	Annelid	ND
Leech (*Placobdella ornate*)	Annelid	NQ
Water flea (*Daphnia magna*)	Arthropoda	ND
Mystery snail (*Pomacea bridgesi*)	Mollusca	ND
Black planaria (*Phagocata gracilis*)	Platyhelminthes	0.99
Brown planaria (*Fugesia tigrine*)	Platyhelminthes	NQ
Caenorhabditis elegans	Nematoda	28.64
Water bear (*Hypsibius sp*.)	Tardigrada	6.63

3.3.4. Effects of Cold, Salinity, and Water Stress on Malylglutamate

We postulated that malylglutamate could serve as an osmolyte, malate/glutamate store, and contribute to charge balance in earthworms.[13] Earthworms (*E. fetida*) were exposed in soil to cold, salinity, and water stress to probe their effects on malylglutamate concentrations in earthworm extracts using ^{1}H NMR. Univariate results are summarized in Tables 3.1 – 3.6, showing the mean, standard deviation, *p*-values and Glass' Δ (i.e., effect size), where $|\Delta| \geq 0.8$ are considered strong and $|\Delta| \geq 1.3$ are considered very strong effect sizes.[10-11] To test if malylglutamate is an osmoprotectant, earthworms were exposed to 0, 10, 40, and 160 mM NaCl, which was introduced in the water used to wet the soil. Soils

with high salt concentrations have an electrical conductivity greater than or equal to 4 dS/m which is equivalent to 40 mM.[20] Malylglutamate, glutamate, and malate levels (Figure 3.13) were not significantly different in the earthworm extracts under the salinity conditions tested. Comprehensive metabolic profiling was not conducted in this study, but of the resonances quantified, alanine, betaine, fumarate, glucose, and α-ketoglutarate decreased with increasing salt content (Table 3.2 – 3.3).

Malylglutamate and malate levels (Figure 3.14) of earthworms at room temperature (RT) were not significantly different to earthworms at 4°C, suggesting that malylglutamate does not serve as a cryoprotectant. Previously, the metabolic impacts of freezing (-2°C) were probed in several earthworm species: *Apporrectodea caliginosa, A. icterica, A. longa, Dendrobaena octaedra*, and *Dendrodrilus rubidus*.[21] Increases in alanine, glucose, lactate, and succinate were observed, along with conversion of adenosine to inosine. Herein, the only statistically significant changes observed at 4°C were increases in glutamate (p = 0.046, Δ = 4.0) and α-ketoglutarate (p = 0.046, Δ = 3.7) (Table 3.4). Lastly, earthworms were exposed to varying water:soil ratios (1:1 (dry), 2:1 (control), and 3:1 (wet)) to study the effect of soil moisture content on earthworm metabolism. Although the levels of malylglutamate, malate, and glutamate did not significantly change under dry conditions (Figure 3.14), decreased malylglutamate levels (p = 0.021, Δ = -1.0) and increased glutamate levels (p = 0.000, Δ = 4.2) were observed comparing 2:1 (control) and 3:1 (wet) groups (Table 3.5 – 3.6). Although malate levels did not change significantly, an increasing trend was observed in the box plot (Figure 3.15). A decrease in malylglutamate and trimethyllysine was observed comparing the same groups (Table 3.5 – 3.6) which may

support the hypothesis that malylglutamate provides charge balance to trimethyllysine and other betaine analogs.[13] Furthermore, changes in TCA cycle and related metabolites (e.g. alanine, glucose, α-ketoglutarate, succinate) suggest a perturbation in energy cycling (Table 3.5-3.6). The significant increase in glutamate without a proportional significant increase in malate (Figure 3.15) may indicate that malylglutamate was hydrolyzed to incorporate malate into the TCA cycle and glutamate accumulated as a byproduct. This supports the hypothesis that malylglutamate is a store for malate and glutamate under stress, and further exploration is needed to understand the relationship between malylglutamate and the TCA cycle.

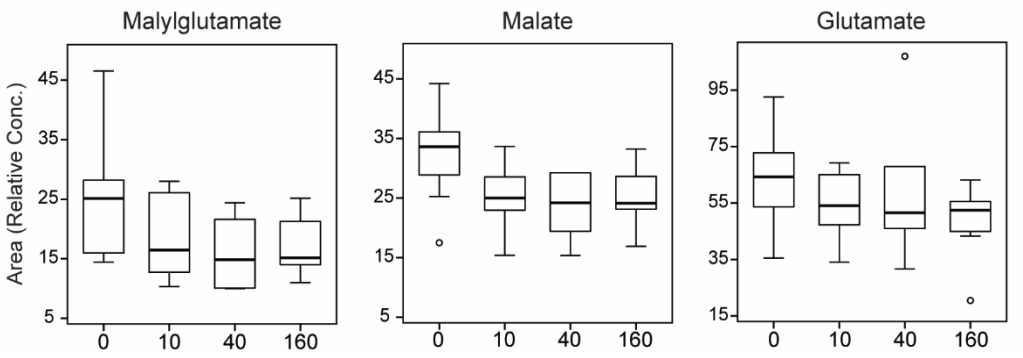


Figure 3.13. Box and whisker plots describing the impact of salinity on malylglutamate, malate, and glutamate levels.

Table 3.2. Summary of ANOVA results with Tukey's HSD and Glass' Δ of statistically significant ($P = 0.05$) metabolic changes among earthworms exposed to 0, 10, 40, and 160 mM NaCl in soil.

mM NaCl	0-10		0-40		0-160		10-160	
	p	Δ	*p*	Δ	*p*	Δ	*p*	Δ
Alanine	0.001	-2.4	0.000	-3.2	0.000	-4.0		
Betaine	0.092	-1.8	0.002	-3.4	0.000	-4.1	0.0	-1.1
Fumarate	0.002	-1.6	0.000	-2.2	0.000	-2.2		
Glucose-x			0.004	-4.1	0.000	-5.0	0.0	-1.3
α-Ketoglutarate					0.046	-1.1	0.1	-1.0

Table 3.3. Summary of mean and standard deviations of statistically significant ($P = 0.05$) metabolic changes among earthworms exposed to 0, 10, 40, and 160 mM NaCl in soil.

mM NaCl	0		10		40		160	
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Alanine	158.43	22.36	103.78	35.21	86.04	31.33	69.08	17.07
Betaine	207.80	26.69	159.13	52.00	117.06	61.07	99.39	29.44
Fumarate	7.15	1.87	4.09	1.48	3.08	2.15	3.04	0.49
Glucose-x	96.34	11.08	80.25	29.12	50.62	31.93	41.23	10.68
α-Ketoglutarate	9.91	4.10	7.13	1.87	7.60	4.60	5.31	2.38

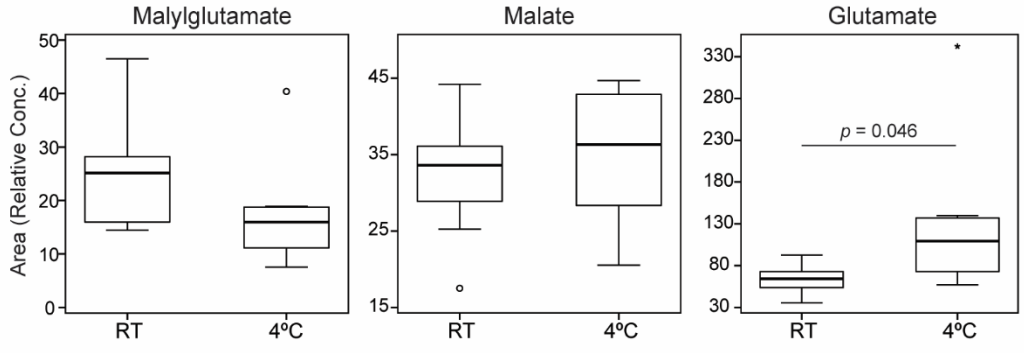


Figure 3.14. Box and whisker plots describing the impact of cold stress on malylglutamate, malate, and glutamate levels.

Table 3.4. Summary of t-test results ($P = 0.05$) of metabolic changes between earthworms exposed to room temperature and 4°C. Means, standard deviation, significance level, and Glass' Δ are reported.

	RT		Cold		RT-Cold	
	$\bar{x}$	σ	$\bar{x}$	σ	*p*	Δ
Glutamate	62.5	16.7	129.7	91.2	0.046	4.0
α-Ketoglutarate	9.9	4.1	25.3	12.9	0.012	3.7

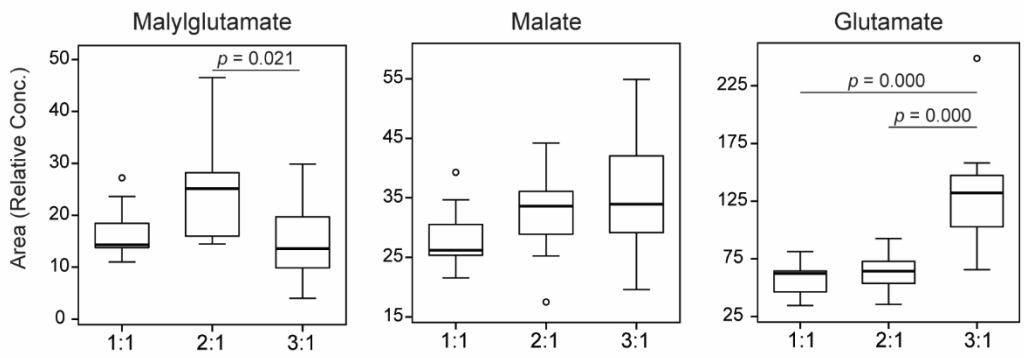


Figure 3.15. Box and whisker plots describing the impact of water:soil ratios on malylglutamate, malate, and glutamate levels.

Table 3.5. Summary of ANOVA results with Tukey's HSD and Glass' Δ of statistically significant (P = 0.05) metabolic changes among earthworms exposed to 1:1 (dry), 2:1 (control) or 3:1 (wet) water:soil conditions.

water:soil	1:1-2:1		1:1-3:1		2:1-3:1	
	p	Δ	*p*	Δ	*p*	Δ
Alanine			0.000	17.2	0.000	15.2
Glucose-x	0.028	1.9			0.060	1.7
Glutamate			0.000	4.8	0.000	4.2
HEFS			0.009	1.8	0.059	0.9
myo-inositol			0.000	6.2	0.000	12.0
α-Ketoglutarate			0.004	-1.4	0.030	-1.3
Malylglutamate					0.021	-1.0
Trimethyllysine	0.010	-2.7			0.050	-1.1
Succinate	0.023	-32.4			0.008	-12.3

Table 3.6. Summary of mean and standard deviations of statistically significant (P = 0.05) metabolic changes among earthworms exposed to 1:1 (dry), 2:1 (control) or 3:1 (wet) water:soil conditions.

water:soil	1:1		2:1		3:1	
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Alanine	145.12	20.58	158.43	22.36	498.43	229.52
Glucose-x	118.74	117.12	96.34	11.08	115.37	21.41
Glutamate	56.58	15.98	62.51	16.72	132.62	50.41
HEFS	652.55	129.77	710.81	201.27	883.17	126.95
myo-inositol	70.84	30.85	72.84	15.83	262.85	119.50
α-Ketoglutarate	11.70	4.98	9.91	4.10	4.56	3.72
Malylglutamate	17.08	5.34	25.31	10.54	14.54	7.77
Trimethyllysine	73.87	28.27	126.79	36.96	86.80	38.43
Succinate	57.34	11.23	80.81	2.15	54.32	17.55

3.3.5. Identification of Malylglutamate as a Chelator

Due to the structural similarity to chelators like *β*-citrylglutamate, it seemed plausible that one of the main functions of malylglutamate in earthworms is to sequester metals which could help explain its high abundance.[13] NMR is a suitable analytical tool for identifying chelators and measuring binding affinity of ligands.[22-23] Mn^{2+} was first selected and $MnCl_2$ was titrated in CF between 0 – 10 μM, where selective peak broadening was observed for the malylglutamate resonances in the ^1H NMR spectra (Figure 3.16), demonstrating that Mn^{2+} binds to malylglutamate. Ca^{2+} and Zn^{2+} were also tested as for binding to malylglutamate between 0 – 10 mM. As observed in Figure 3.17, the malyl resonances of malylglutamate (i.e., HABX) shifted downfield significantly and Hγ shifted slightly downfield with increasing Ca^{2+} concentration. All the malylglutamate resonances labeled, except Hα, significantly shifted downfield with increasing Zn^{2+} concentration and peaks broadened slightly (Figure 3.18). Competition of other CF metabolites for metal complexation is observed, particularly malate. This survey confirms that malylglutamate is indeed a chelator, and further exploration is needed to compare affinity of metals for malylglutamate to identify if it is a selective or general chelator.

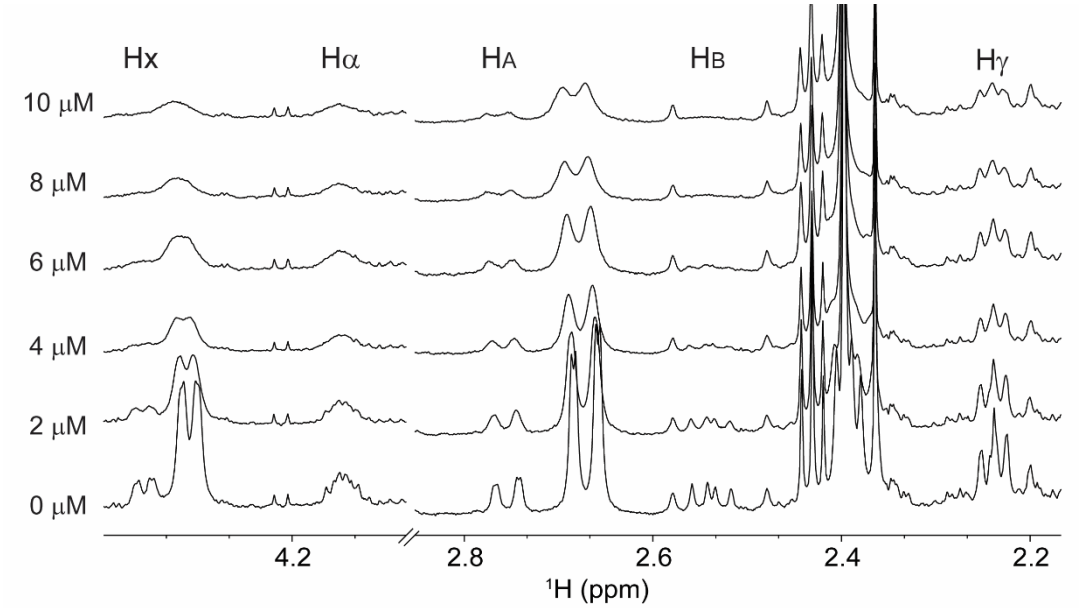


Figure 3.16. ^{1}H NMR survey spectra of CF titrated with 0 – 10 μM Mn^{2+} showing the selective broadening and slight downfield shift of malylglutamate resonances with increasing Mn^{2+}.

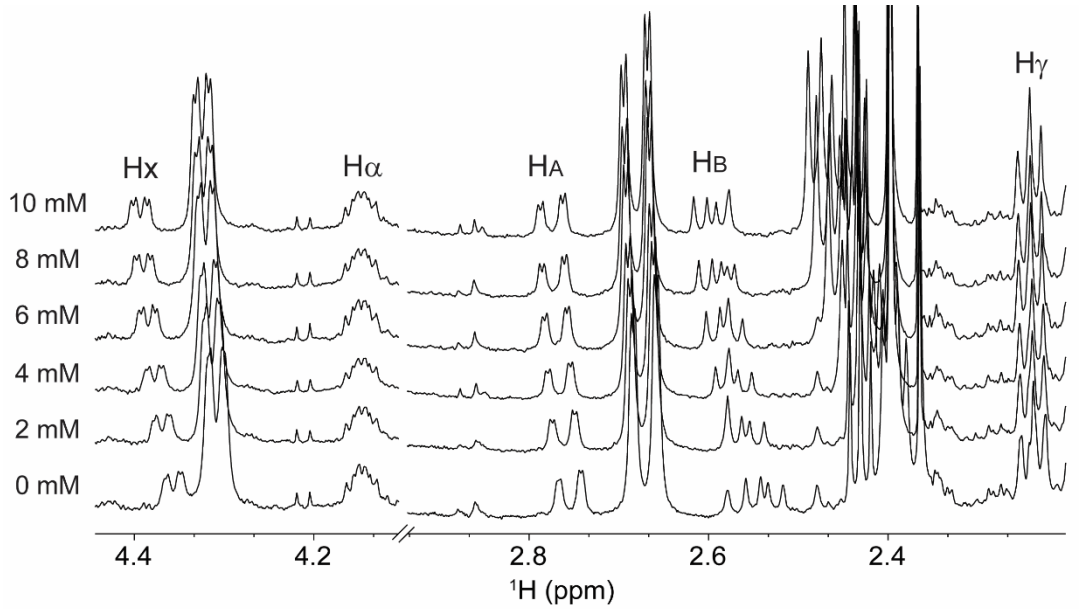


Figure 3.17. ^{1}H NMR survey spectra of CF titrated with 0 – 10 mM Ca^{2+} showing the resonances of malylglutamate significantly shift slightly downfield with increasing Ca^{2+}.

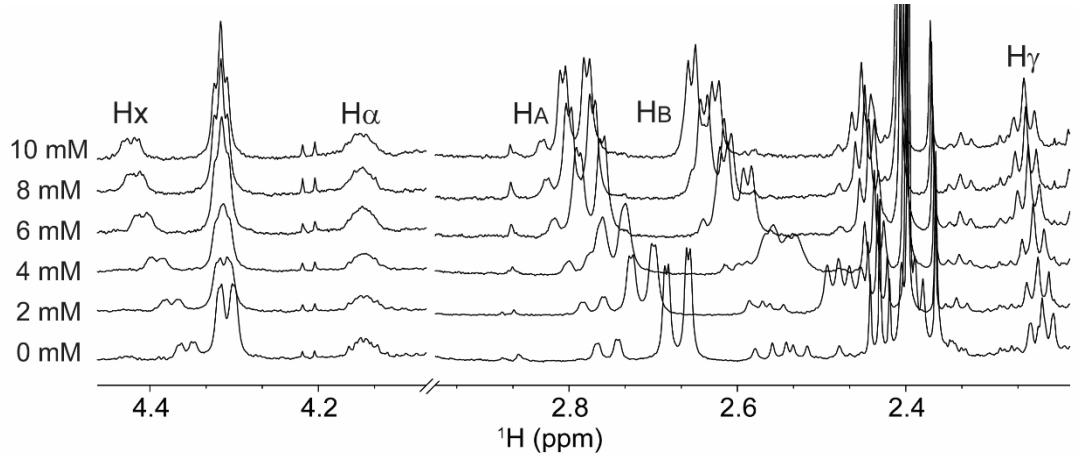


Figure 3.18. ^{1}H NMR survey spectra of CF titrated with 0 – 10 mM Zn^{2+} showing the resonances of malylglutamate significantly shift slightly downfield and broaden slightly with increasing Zn^{2+}.

3.4. Conclusions

The structure elucidation of malylglutamate, a major earthworm metabolite, was a significant outcome of this dissertation. The absolute configuration of the molecule, (–)-*β*-L-malyl-L-glutamate, was elucidated using NMR spectroscopy, high-resolution mass spectrometry, synthesis, chiral chromatography, and optical rotation. Surprisingly, malylglutamate was not previously reported in literature, but due to its simple structure and high abundance, it seemed plausible that it was present in other organisms. LC-MS/MS detected malylglutamate in several species closely related on the phylogenetic tree, including other earthworm species. Malylglutamate was present in μM – mM concentrations in nightcrawlers, black worms, leeches, black planaria, *C. elegans*, and water bears, but it was not detected in white worms, brown planaria, mystery snails, or

Daphnia magna. Further exploration should be conducted to screen for malylglutamate in other organisms, including humans, to determine its ubiquity and if it is one of the many known unknowns in metabolomics datasets.

Malylglutamate has been detected at approximately 7 μg/mg in earthworms and we postulated that with such a high abundance it must serve important biological functions. We hypothesized that malylglutamate was an anionic osmolyte; however, malylglutamate levels did not significantly change under salinity stress in earthworms and therefore it does not appear to be osmoprotective. Malylglutamate also does not appear to serve as a cryoprotectant since its levels did not significantly change under cold stress. Malylglutamate was also thought to be a store for malate and glutamate under increased energy demand, which is supported by an observed significant decrease in malylglutamate and increase in glutamate in earthworms exposed to high water content. Malate levels did not significantly change, but since other TCA cycle and related metabolites were perturbed, we suggest that malylglutamate was hydrolyzed to free up malate for the TCA cycle and glutamate accumulated as a byproduct. Additionally, levels of trimethyllysine decreased along with malylglutamate under high water content which potentially supports our hypothesis that malylglutamate provides charge balance for trimethyllysine and betaine analogs. Lastly, we hypothesized that malylglutamate was a chelator and we demonstrated that Ca^{2+}, Mn^{2+}, and Zn^{2+} were complexed using ^{1}H NMR. Further studies are needed to elucidate the biological functions of malylglutamate, including its relationship with the TCA cycle, and the specificity of metal chelation by malylglutamate should be explored to

evaluate if it is designed to protect earthworms from metal burdens or to sequester micronutrients.

3.5. References

1. Madalinski, G.; Godat, E.; Alves, S.; Lesage, D.; Genin, E.; Levi, P.; Labarre, J.; Tabet, J.-C.; Ezan, E.; Junot, C., Direct introduction of biological samples into a ltq-orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal. Chem.* **2008,** *80* (9), 3291-3303.

2. Chan, T. M.; Kong, J.; McNamara, P.; Wong, J. K., Syntheses of potential degradation products of phenylephrine in otc products. *Synth. Commun.* **2008,** *38* (13), 2252-2260.

3. Du, Z.; Lu, Y.; Dai, X.; Zhang-Negrerie, D.; Gao, Q., The discovery of a facile access to the synthesis of nsaid dendritic prodrugs. *J. Chem. Res.* **2013,** *37* (3), 177-180.

4. Schobert, R.; Jagusch, C., An efficient synthesis of carlosic acid and other 5-carboxymethyltetronates from malates. *Synthesis* **2005,** *2005* (14), 2421-2425.

5. Dutton, F. E.; Lee, B. H., Epsilon-lactam analogs of the anthelmintic cyclodepsipeptide pf1022a. *Tetrahedron Lett.* **1998,** *39* (30), 5313-5316.

6. Bax, A.; Subramanian, S., Sensitivity-enhanced two-dimensional heteronuclear shift correlation nmr spectroscopy. *J. Magn. Reson.* **1986,** *67* (3), 565-569.

7. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-968.

8. Bligh, E. G.; Dyer, W. J., A rapid method of total lipid extraction and purification. *Can. J. Biochem. Phys.* **1959,** *37* (8), 911-917.

9. Levene, H., Robust tests for equality of variances. In *Contributions to probability and statistics: Essays in honor of harold hotelling*, Olkin, I., Ed. Stanford University Press: 1960; pp 278-292.

10. Ialongo, C., Understanding the effect size and its measures. *Biochem. Med.* **2016,** *26* (2), 150-163.

11. Rosenthal, J. A., Qualitative descriptors of strength of association and effect size. *J. Soc. Serv. Res.* **1996,** *21* (4), 37-59.

12. Collard, F.; Stroobant, V.; Lamosa, P.; Kapanda, C. N.; Lambert, D. M.; Muccioli, G. G.; Poupaert, J. H.; Opperdoes, F.; Van Schaftingen, E., Molecular identification of n-acetylaspartylglutamate synthase and β-citrylglutamate synthase. *J. Bio. Chem.* **2010,** *285* (39), 29826-29833.

13. Griffith, C. M.; Williams, P. B.; Tinoco, L. W.; Dinges, M. M.; Wang, Y.; Larive, C. K., 1h nmr metabolic profiling of earthworm (eisenia fetida) coelomic fluid, coelomocytes, and tissue: Identification of a new metabolite – malylglutamate. *J. Proteome Res.* **2017,** *16* (9), 3407-3417.

14. Collard, F.; Vertommen, D.; Constantinescu, S.; Buts, L.; Van Schaftingen, E., Molecular identification of beta-citrylglutamate hydrolase as glutamate carboxypeptidase 3. *J. Bio. Chem.* **2011,** *286* (44), 38220-38230.

15. Ciulla, R.; Clougherty, C.; Belay, N.; Krishnan, S.; Zhou, C.; Byrd, D.; Roberts, M. F., Halotolerance of methanobacterium thermoautotrophicum delta h and marburg. *J. Bacteriol.* **1994,** *176* (11), 3177-3187.

16. Hamada-Kanazawa, M.; Kouda, M.; Odani, A.; Matsuyama, K.; Kanazawa, K.; Hasegawa, T.; Narahara, M.; Miyake, M., Beta-citryl-l-glutamate is an endogenous iron chelator that occurs naturally in the developing brain. *Biol. Pharm. Bull.* **2010,** *33* (5), 729-737.

17. Narahara, M.; Hamada-Kanazawa, M.; Kouda, M.; Odani, A.; Miyake, M., Superoxide scavenging and xanthine oxidase inhibiting activities of copper s-citryl-l-glutamate complex. *Biol. Pharm. Bull.* **2010,** *33* (12), 1938-1943.

18. Hamada-Kanazawa, M.; Narahara, M.; Takano, M.; Min, K. S.; Tanaka, K.; Miyake, M., Beta-citryl-l-glutamate acts as an iron carrier to activate aconitase activity. *Biol. Pharm. Bull.* **2011,** *34* (9), 1455-1464.

19. Hamada-Kanazawa, M.; Narahara, M.; Takano, M.; Min, K. S.; Tanaka, K.; Miyake, M., Nitric oxide promotes survival of cerebral cortex neurons with simultaneous addition of fe(ii)(beta-citryl-l-glutamate) complex in primary culture. *Biol. Pharm. Bull.* **2013,** *36* (7), 1068-1079.

20. Munns, R., Genes and salt tolerance: Bringing them together. *New Phytol.* **2005,** *167* (3), 645-663.

21. Bundy, J. G.; Ramløv, H.; Holmstrup, M., Multivariate metabolic profiling using 1h nuclear magnetic resonance spectroscopy of freeze-tolerant and freeze-intolerant earthworms exposed to frost. *Cryoletters* **2003,** *24* (6), 347-358.

22. Kadima, W.; Rabenstein, D. L., Nuclear magnetic resonance studies of the solution chemistry of metal complexes. 26. Mixed ligand complexes of cadmium, nitrilotriacetic acid, glutathione, and related ligands. *J. Inorg. Biochem.* **1990,** *38* (4), 277-288.

23. Chang, J.-Y.; Carollo, K. D.; Lin, S.-C.; Wu, Y.-Y.; Tzou, D.-L. M., Nmr investigation of magnesium chelation and cation-induced signal shift effect of testosterone. *Steroids* **2016,** *115*, 18-25.

CHAPTER FOUR

Metabolic profiling of chloroacetanilide herbicides in earthworm coelomic fluid using ^{1}H NMR and GC-MS

Based in part on a paper published in the *Journal of Proteome Research*

J. Proteome Res. 2018, *17* (8), 2611-2622

Acknowledgements: I would like to thank Drs. Melissa Morgan, Meredith Dinges, and Caroline Mathon for assisting with the bench and statistical analyses conducted in this chapter.

Abstract

Earthworms (*Eisenia fetida*) are vital members of the soil environment. Because of their sensitivity to many contaminants, monitoring earthworm metabolism may be a useful indicator of environmental stressors. Here, metabolic profiles of exposure to five chloroacetanilide herbicides and one enantiomer (acetochlor, alachlor, butachlor, racemic metolachlor, S-metolachlor, and propachlor) are observed in earthworm coelomic fluid using proton nuclear magnetic resonance spectroscopy (NMR) and gas chromatography-mass spectrometry (GC-MS). Multiblocked-orthogonal partial least squares-discriminant analysis (MB-OPLS-DA) and univariate analysis were used to identify metabolic perturbations in carnitine biosynthesis, carbohydrate metabolism, lipid metabolism, nitrogen metabolism, and the TCA cycle. Intriguingly, stereospecific metabolic responses were observed between racemic metolachlor and S-metolachlor exposed worms. These findings support the utility of coelomic fluid in monitoring metabolic perturbations induced

by chloroacetanilide herbicides in nontarget organisms and reveal specificity in the metabolic impacts of herbicide analogues in earthworms.

4.1. Introduction

The relationship between molecular structure and metabolic impact is seldom studied, although understanding structure-activity relationships (SARs) can aid the development and prediction of adverse outcome pathways (AOPs) and exposure biomarkers. This study aimed to examine how exposure to members of the structurally analogous chloroacetanilide herbicides affected the earthworm (*Eisenia fetida*) coelomic fluid (CF) metabolite profiles to gain insights into how structure influences metabolic response in a nontarget organism. Earthworms are sensitive to contaminants, found in most terrestrial ecosystems, and are essential ecological engineers responsible for degrading plant matter and providing soil aeration; thus they are a clear choice for environmental monitoring.[1] Many studies have used metabolomics to examine the impact of heavy metals, industrial contaminants, and pesticides on the earthworm metabolome.[2-3] Compared to the profile obtained using whole-worm extracts, we hypothesize that metabolic profiling of earthworm CF can provide a complementary record of the impact of xenobiotic compound exposure due to the central role of CF in earthworm homeostasis.

Chloroacetanilide herbicides, such as acetochlor, alachlor, butachlor, racemic metolachlor, S-metolachlor, and propachlor (Fig. 4.1), are a class of broad-spectrum, pre-emergent herbicides used to target broadleaf weeds and grasses. The chloroacetanilide

mechanism of action (MOA) is not completely understood, but they are known to target the gibberellin pathway, lipid biosynthesis, and elongation.[4-5] Additionally, these herbicides can produce toxic effects in nontarget organisms; for example, the U.S. Environmental Protection Agency set a maximum contaminant level of 2 ppb for alachlor in drinking water due to its kidney, spleen, eye, and liver toxicity, along with increased risk of cancer.[6]

Chloroacetanilides are mobile in the environment and are frequently detected in water systems, especially after transformation to their hydrophilic ethane sulfonic acid (ESA) and oxanilic acid (OA) derivatives.[7-9] The application of chloroacetanilide herbicides is limited to S-metolachlor in the European Union,[10] while use in the U.S. varies by state but several of these herbicides are still extensively used in the U.S. Midwest, especially in corn and soybean cultivation. In 2015, the U.S. Geological Survey estimated that more than 22×10^6 kg S-metolachlor and 18×10^6 kg of acetochlor were applied in the U.S.[11] Less than 4.5×10^6 kg of alachlor and racemic metolachlor were estimated, while the application of butachlor and propachlor were not reported.[11] A recent study (2010-2012) detected alachlor in sediments samples from the great lakes at median concentrations between 0.001 and 0.18 ng/g-dry weight.[12] Metolachlor, metolachlor ESA, metolachlor OA, alachlor ESA, and alachlor were the 8th, 18th, 52nd, 150th, and 293rd, respectively, most frequently detected anthropogenic compounds in a nationwide US study of stream surface water, sampled between 2012-2014.[13] Thus with their toxicity, mobility, and potential to leach into groundwater and runoff into surface waters, it is critical to monitor the nontarget impacts of chloroacetanilide herbicides on soil ecosystems. In addition, a unique feature of

this study is that the closely related structures of the compounds examined provide the ability to probe metabolic SARs in an important nontarget organism.

Herein, ^{1}H NMR and GC-MS were used to profile the metabolomes of earthworm coelomic fluid after exposure to chloroacetanilide herbicide to probe mode of action and compare metabolic impacts. Combining datasets acquired using multiple analytical platforms simplifies the interpretation of metabolic profiles, as is evident in literature and from the results presented herein to assess the impact of chloroacetanilide herbicide exposure in earthworms.[14-17]

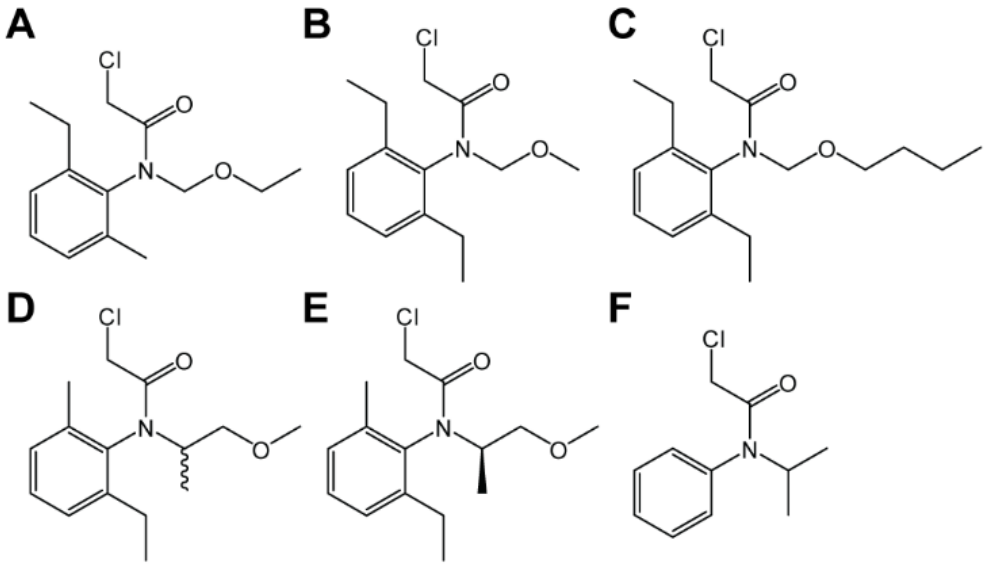


Figure 4.1. Chemical structures of chloroacetanilide herbicides: A) acetochlor, B) alachlor, C) butachlor, D) racemic metolachlor, E) S-metolachlor, and F) propachlor.

4.2. Experimental Procedures

4.2.1. Exposure and Herbicides

Earthworm culturing was carried out as previously described.[18] Exposures were conducted following the Organization for Economic Co-operation and Development (OECD) acute toxicity earthworm guidelines for the filter paper test.[19] An amber vial was lined with a 9 cm VWR 413 qualitative filter paper (Randor, PA) and 1 mL of herbicide in acetone solution was applied to the paper one day prior to the start of the experiment and stored in the freezer overnight. The vials and paper were dried under a nitrogen stream prior to moistening with 1 mL of ultrapure water (EMD Millipore Simplicity, Billerica, MA). An adult earthworm with a fully developed clitellum that had been depurated for 24 h was added to each vial. A Kimwipe® was wrapped around the opening with a rubber band to allow ventilation and the earthworm was left on the filter paper for two days in the dark.

Pestanal® (Sigma-Aldrich, St. Louis, MO) acetochlor, alachlor, butachlor, metolachlor, S-metolachlor, and propachlor were chosen to explore the metabolic impacts of the chloroacetanilide herbicides. Similar LC_{50} concentrations using the filter paper test have been reported in literature for chloroacetanilide herbicide exposure in earthworms, with acetochlor yielding a LC_{50} dose of 24.02 $\mu g/cm^2$, 20.63 $\mu g/cm^2$ for S-metolachlor, and 19.23 $\mu g/cm^2$ for pretilachlor (not used in this study).[20] We chose the acetochlor LC_{50} as the base dose for these experiments since it was the highest of the three LC_{50} doses reported. Each compound was dosed in 6 replicates of 50% (12.01 $\mu g/cm^2$), 25% (6.01 $\mu g/cm^2$), and 10% (2.40 $\mu g/cm^2$) the reported acetochlor filter test LC_{50} of 24.02 $\mu g/cm^2$.[20]

Control worms were treated by exposure to a filter paper dosed with 1 mL acetone and prepared identically to the herbicide dosed vials described above. Experiments were staggered so that half the samples were started on day one (i.e., 60 worms: 6 control worms and 54 dosed worms for alachlor, metolachlor, S-metolachlor) and the second set (i.e., 60 worms: 6 control worms and 54 dosed worms for acetochlor, butachlor, and propachlor) was started on day two. The controls from both sets were grouped together in the analyses, resulting in a total of 12 control worms and thus, N = 120 worms. Five worms died over the course of the experiment (acetochlor – 6.01 μg/cm^2; alachlor – 12.01 μg/cm^2; butachlor – 2.40 μg/cm^2; S-metolachlor – 12.01 μg/cm^2; and propachlor – 6.01 μg/cm^2).

4.2.2. Coelomic Fluid (CF) Extrusion

Following exposure, each earthworm was transferred to a 350 mm Falcon Petri dish and CF was extruded as described in Chapter 2, Section 2.2.2. Samples were centrifuged for 20 min at 16100 *g* to pellet the coelomocytes and biosolids and 100 μL of the supernatant transferred to a 350 μL glass flat-bottom GC insert. The remaining supernatant was transferred to a fresh Eppendorf tube for NMR analysis. The NMR and GC-MS samples were evaporated at ambient room temperature overnight using a Savant SC110 speedvac equipped with a refrigerator vapor trap (RVT400). The dried samples were stored at -80°C until analysis.

4.2.3. Metabolite Measurements using ^{1}H NMR

The dried samples for NMR were reconstituted with 500 μL of 50 mM phosphate buffer (pD 7.45) in D_2O (D, 99.9%) (Cambridge Isotope Laboratories, Tewksbury, MA) containing 0.5 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonic acid-d_6 (DSS-d_6) and 0.2 mM ethylenediaminetetraacetic acid-d_{16} (Cambridge Isotope Laboratories, Tewksbury, MA) and vortexed. To remove glycerol, 3K Amicon Ultra 0.5 mL centrifugal filters (EMD Millipore, Billerica, MA) were washed over 60 h by stirring in approximately 2 L ultrapure water, with the water replaced 5 times. The filter was rinsed a final time with 500 μL of D_2O by centrifugation at 16100 *g* for 15 min to reduce the signal due to protonated water in the filtrate. Each sample was passed through a washed centrifugal filter at 16100 *g* for 30 min to remove macromolecules. A 400 μL aliquot of the filtered sample was transferred to a 5 mm NE-UL5-7" NMR tube (New Era Enterprises, Vineland, NJ), 200 μL buffer was added, and the sample was homogenized.

One-dimensional ^{1}H NMR spectra were acquired using a Bruker Avance III NMR spectrometer operating at 700.23 MHz and equipped with a TCI CryoProbe. Water suppression was conducted using 1D nuclear Overhauser effect spectroscopy (NOESY) with presaturation (noesypr1d) during the 120 ms mixing time and 2 s relaxation delay. Spectra were acquired at 25°C with 256 free induction decays (FIDs) coadded, 16 dummy scans, 2.03 s acquisition period for a total experiment time of 18 min and 58 s. A spectral width of 11.5169 ppm was used with 32768 complex data points acquired using digital quadrature detection.

4.2.4. ^{1}H NMR Data Preprocessing

Data preprocessing was initially conducted using Bruker TopSpin 3.2 (Billerica, MA) for initial phasing and chemical shift referencing to DSS-d_6 (0 ppm) prior to processing with MestReNova 11 (Santiago de Compostela, Spain). FIDs were apodized by multiplication with an exponential function equivalent to 1 Hz line broadening, zero-filled to 131072 points, drift corrected by 5%, and baseline corrected using a Whittaker Smoother function set with a medium filter of 2.36 Hz and smooth factor of 32768. Peak fitting was conducted using a generalized Lorentzian peak shape, a lower width constraint of 0 Hz, an upper width constraint of 30 Hz, position constraint within ± 5%, maximum number of fine iterations of 100, and local minima filter of 0. The peak fitting results were exported to Excel (Microsoft Corporation, Redmond, WA) and assembled into a single spreadsheet. The results of each spectrum were normalized to the sum of the total area between 0.8-9.0 ppm, excluding the HOD peak between 4.6-5.2 ppm. Resonance assignments were previously described,[18] and are summarized in Chapter 2.

4.2.5. Metabolite Measurements using GC-MS

GC-MS aliquots were derivatized prior to analysis by adding 20 μL of 20 mg/mL methoxyamine (Sigma-Aldrich, St. Louis, MO) in pyridine (Thermo Scientific, Bellefonte, PA) and mixing at 800 rpm for 90 min at 37°C.[21] A 2 μL aliquot of a fatty acid methyl ester (FAMES) standard containing 0.8 mg/mL C8, C9, C10, C12, C14, and C16 and 0.4 mg/mL C18, C20, C22, C24, C26, C28, and C30 was added to each sample as a retention time reference, along with 90 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide

(MSTFA) containing 1% chlorotrimethylsilane (TMCS) (Thermo Scientific, Bellefonte, PA). The silylation reaction was carried out for 30 min at 37°C. Quality control (QC) samples were prepared from a 5-worm pooled CF sample in 1 mL, 20 μL divided to GC inserts, and prepared identically. A fresh QC sample was prepared and injected daily to evaluate consistency between days, which was observed. GC-MS analysis was conducted using previously described methods.[21] Briefly, samples were injected in pulsed splitless mode on an Agilent J&W DB-5MS UI 30 m x 0.25 mm x 0.25 μm column (Santa Clara, CA) using an Agilent 7890A gas chromatograph coupled to a Waters GCT Premier mass spectrometer. The oven temperature was initially set at 60°C and ramped to 320°C at 10°C/min, followed by a 5 min hold at the final temperature. The injector, transfer line, and source were maintained at 230°C, 320°C, and 220°C, respectively, and the liner was changed after every 25 injections.

4.2.6. GC-MS Data Preprocessing

Data preprocessing was performed using MarkerLynx XS v4.1 (Waters Corporation, Milford, MA) software using parameters previously reported, with adjustments only in the initial retention time of 5.50 min and final retention time of 32.00 min.[21] A 73.5 Da low-mass cut off and 600 Da high-mass cut off with a 0.10 Da mass accuracy window was used to detect deisotoped peaks without smoothing and the peak width at 5% was overestimated by 2 s. A mass retention window of 0.1 Da/min and 25 count marker intensity threshold was selected. A peak-to-peak baseline noise value of 1 was used alongside a noise elimination level of 3. Compounds were identified using the

Golm Metabolome Database and NIST 2017 Mass Spectral Library in Automated Mass Spectral Deconvolution and Identification System (AMDIS) software.[22] The analysis of authentic standards and retention indices calculated using the FAMES standards were used to confirm assignments. Results were exported to Excel and a single *m/z*-retention time pair that was not representative of a silylation fragment was used to quantify each metabolite. The peak areas of metabolites with multiple peaks (i.e., degrees of silylation) were summed. The results for each sample were normalized to the sum of the total area, excluding areas before 6.0 min, 7.60-7.67 min and 8.39-8.42 min (silylates), 9.04-9.09 min (phosphoric acid), 18.01-18.08 min (hexadecanoate), 19.81-19.87 min (octadecanoate), and FAMES.

4.2.7. Chemometric Analyses

Multiblocked multivariate analysis was conducted in SIMCA 14.1 (Sartorius Stedim Biotech, Malmö, Sweden). Homoscedasticity was assured for all variables in each model and variables flagged as heteroscedastic by the software were transformed logarithmically to correct skewness. Unit variance scaling and centering was applied to the MB dataset and each block weight was set to 1/sqrt. Multiblocked-orthogonal partial least squares-discriminant analysis (MB-OPLS-DA) was used to evaluate grouping and cross validate-analysis of variance (CV-ANOVA) with $p \leq 0.05$ used to confirm fitting.[23] Multiblocked-partial least squares-hierarchical cluster analysis (MB-PLS-HCA) was also performed in SIMCA using multiblocked-partial lease squares (MB-PLS) dimension

reduction of all samples and grouped as function of treatment (i.e., not considering dose). Distance was calculated using Ward's method and the dendrogram was sorted by size.

Normalized data were imported from Excel into IBM SPSS Statistics v24 (Armonk, NY) for univariate analysis. Data were divided into control-herbicide pairs prior to univariate analysis. Levene's test was used to confirm homoscedasticity among independent variables prior to analysis of variance (ANOVA).[24] Variables that violated the normality assumption were transformed using square, square root, or natural log depending on the variable's skewness. Pairwise comparisons were conducted using Tukey's HSD with a p-value ≤ 0.10. Effect size for metabolites with statistically significant changes were calculated using Glass' Δ.[25]

4.3. Results and Discussion

The first step in assessing the metabolic impact of chloroacetanilide herbicides on the *E. fetida* CF metabolome was to identify the metabolite signatures in the ^{1}H NMR spectra and GC-MS chromatograms. We detected 47 metabolites by NMR (Figure 4.2), 44 by GC-MS (Figure 4.3), and 22 with both methods (Figure 4.4). The earthworm ^{1}H NMR CF metabolome has been well characterized.[18, 26-27] Though GC-MS has been widely used for the characterization of earthworm extracts,[28-33] to our knowledge, the metabolite profiling of CF using GC-MS has not been reported prior to this study. Multiple amino and organic acids were detected using both platforms (Figure 4.4); however, adipate, 2-aminoglutarate, azelate, citrate, cholesterol, glutarate, glycerol-3-phosphate, lysine, 2-methylglutarate, ornithine, proline, serine, suberate, and urea were detected only by GC-

MS (Figure 4.3, Table 4.1). ^{1}H NMR was superior in detecting betaine analogs like *N,N*-dimethylhistidine (DMH), *Nε,Nε,Nε*-trimethyllysine (TML), and *Nδ,Nδ,Nδ*-trimethylornithine (TMO); however, fatty acid resonances were absent or difficult to quantify in our ^{1}H NMR spectra. GC-MS effectively offered insights into fatty acids, such as dodecanoate, eicosanoate, nonanoate, tetradecanoate, and triacontanoate (Figure 4.3, Table 4.1). Phosphoric acid, hexadecanoate, and octadecanoate were also detected in our samples, but were not considered in the analysis of herbicide impact due their high variability.

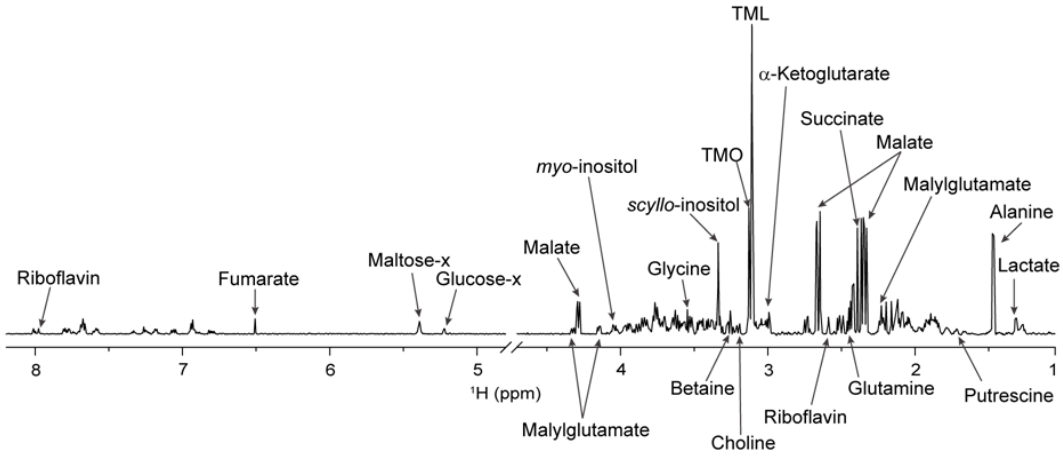


Figure 4.2. Representative ^{1}H NMR spectrum of a control earthworm coelomic fluid sample. A subset of detected metabolites is annotated, with a full description of resonance assignments described in Griffith et al.[11] and discussed in Chapter 2. *TML = *Nε,Nε,Nε*-trimethyllysine; TMO= *Nδ,Nδ,Nδ*-trimethylornithine.

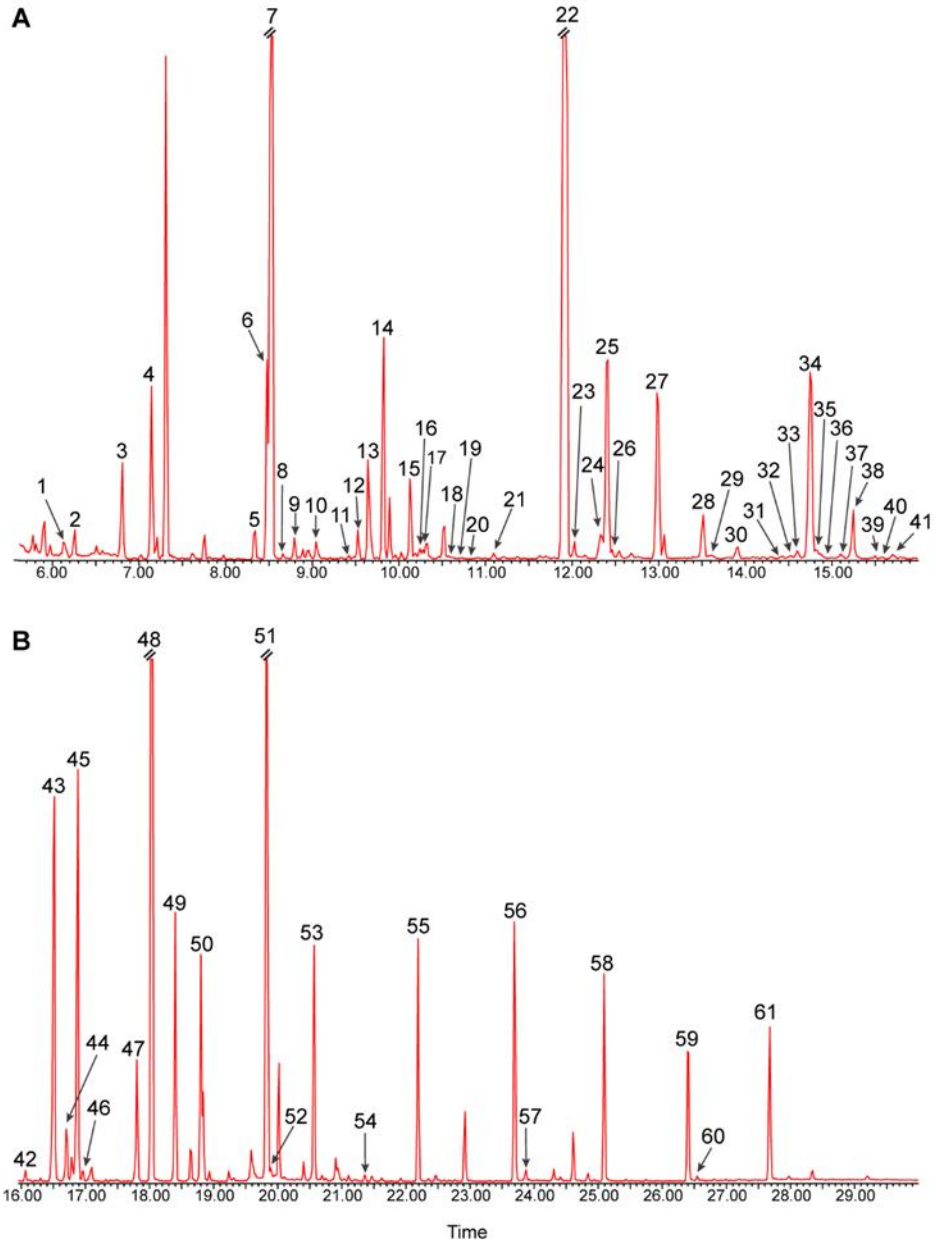


Figure 4.3. Representative GC-MS total ion chromatogram (TIC) between A) 5.00 – 16.00 min and B) 16.00 – 30 min of a control earthworm coelomic fluid sample. Metabolites are annotated with numbers in Table 4.1.

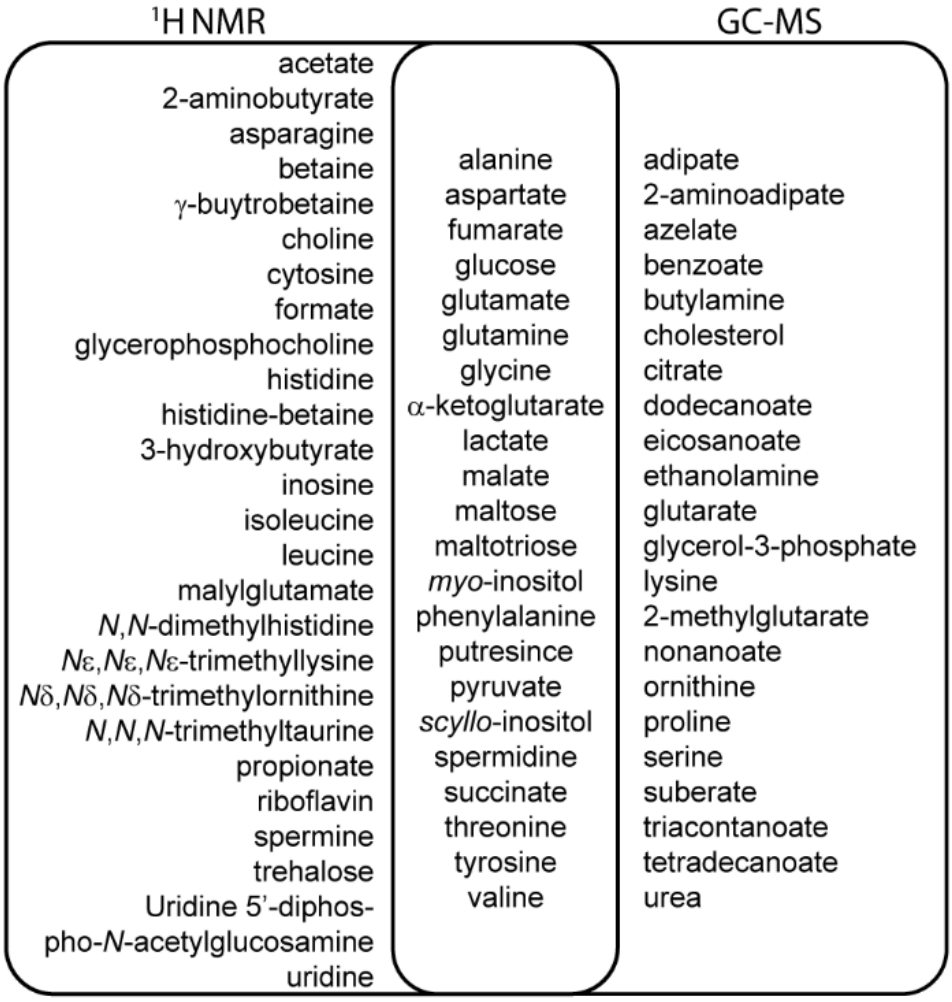


Figure 4.4. Venn diagram comparing metabolites detected with ^{1}H NMR (left), GC-MS (right), or both (center). Phosphoric acid, hexadecanoate, and octadecanoate were also detected in our GC-MS samples, but were not considered in the analysis of herbicide impact due their high variability.

Table 4.1. GC-MS metabolite assignment with retention time, calculated retention index, reported retention index, and *m/z* used for quantitation. Retention Index (RI) is reported from NIST 2017.

#	Metabolite	Retention Time	Retention Index (Calculated)	Retention Index (NIST)	*m/z* for quantitation
1	Pyruvate (1TMS)	6.126	1064	1037*	174.08
2	Lactate (2TMS)	6.248	1072	1066	117.09
3	Alanine (2TMS)	6.791	1109	1100	116.10
4	FAMES: C8	7.125	1132	1126	-----
5	Butylamine (2TMS)	8.312	1218	1123	130.12
6	FAMES: C9	8.451	1124	1225	-----
7	Ethanolamine (3TMS)	8.511	1130	1242	174.13
8	Urea (2TMS)	8.607	1238	1249	-----
9	Benzoate (1TMS)	8.778	1251	1249	179.06
10	Phosphate (3TMS)	9.044	1294	1292	-----
11	Proline (2TMS)	9.402	1296	1305	142.12
12	Glycine (3TMS)	9.518	1304	1314	174.14
13	Succinate (2TMS)	9.636	1313	1321	75.04
14	FAMES: C10	9.813	1325	1325	-----
15	Fumarate(2TMS)	10.127	1350	1353	245.10
16	Serine (3TMS)	10.236	1357	1368	204.15
17	Nonanoate (1TMS)	10.247	1361	1355	117.04
18	Threonine (3TMS)	10.578	1383	1367	218.17
19	Glutarate (2TMS)	10.876	1406	1409	75.03
20	2-Methylglutarate (2TMS)	10.992	1414	1423	239.20
21	Aspartate(2TMS)	11.092	1422	1420	75.03
22	Malate (3TMS)	11.916	1484	1497	133.03
23	Adipate (2TMS)	12.142	1502	1514	75.03
24	Aspartate (3TMS)	12.309	1513	1522	232.15
25	FAMES: C12	12.403	1521	1526	-----
26	Glutamate (2TMS)	12.448	1525	1527	84.06
27	α-Ketoglutarate	12.985	1571	1587	198.09
28	Glutamate (3TMS)	13.508	1616	1651	246.17
29	Phenylalanine (2TMS)	13.606	1624	1636	91.07
30	Dodecanoic acid (1TMS)	13.905	1651	1655	117.05
31	Suberate (2TMS)	14.424	1696	1707	75.04
32	Lysine (3TMS)	14.513	1703	1718	84.09
33	2-Aminoadipate	14.621	1711	1719	260.18

34	FAMES: C14	14.751	1724	1725	-----
35	Glutamine (4TMS)	14.763	1725	1721*	227.17
36	Putrescine (4TMS)	14.835	1731	1733	174.14
37	Glycerol-3-phospate	15.097	1757	1778	-----
38	Glutamine (3TMS)	15.241	1771	1768	156.11
39	Azelate (2TMS)	15.505	1795	1806	317.20
40	Ornithine	15.707	1815	1812	174.14
41	Citrate (3TMS)	15.711	1815	1845	273.14
42	Tetradecanoate (1TMS)	16.061	1849	1850	285.24
43	Glucose	16.509	1891	1865	205.13
44	Lysine (4TMS)	16.780	1918	1933	174.14
45	FAMES: C16	16.872	1927	1926	-----
46	Tyrosine (3TMS)	16.956	1935	1959	218.13
47	*scyllo*-Inositol (6TMS)	17.798	2019	2027*	318.20
48	Hexadecanoate (1TMS)	18.034	2070	2049	-----
49	*myo*-Inositol (6TMS)	18.397	2078	2129	217.14
50	FAMES: C18	18.796	2117	2128	-----
51	Octadecanoate (1TMS)	19.833	2277	2234	-----
52	Spermidine (5TMS)	19.879	2249	2245	174.14
53	FAMES: C20	20.559	2333	2329	-----
54	Eicosanoate (1TMS)	21.462	2443	2447	117.04
55	FAMES: C22	22.183	2531	2528	-----
56	FAMES: C24	23.688	2737	2728	-----
57	Maltose	23.704	2739	2733	361.22
58	FAMES: C26	25.087	2940	2935	-----
59	FAMES: C28	26.396	3112	3126	-----
60	Cholesterol (1TMS)	26.543	3131	3150	-----
61	FAMES: C30	27.666	3279	3323	-----
62	Triacontanoate (1TMS)	28.338	3366	3426	117.04
63	Maltotriose (11TMS)	29.201	3481	3508*	204.13

*RIs from the Golm Metabolome Database

**---- = not quantified

Chemometric analyses were conducted with multivariate and univariate statistical approaches to elucidate exposure biomarkers and evaluate how the metabolic perturbations arising from exposure varied across this family of structurally similar compounds. Multivariate analysis was conducted using MB-OPLS-DA and CV-ANOVA with a 95% significance level used to evaluate significance of separation. To ascertain the metabolic drivers of the separation, a MB-OPLS-DA loadings heatmap (Figure 4.5A) was constructed by plotting metabolites with $|p(corr)| > 0.4$ and Variable Importance in Projection (VIP) scores > 1. This strategy compares metabolites that contribute more than the average to the separation (i.e., VIP > 1) with loadings scaled as a correlation coefficient (i.e., p(corr)) between the model and original data, which is commonly visualized in volcano plots.[23, 34] Metabolites that met these criteria in control samples were plotted in green in Figure 4.5A and the metabolites associated with exposure are indicated in blue.

Figures 4.5B-G present the score plots for each model, with the segregation component (i.e., t[1]) plotted against the orthogonal component (i.e., t_o[1]) which represents class variability that is unrelated to class discrimination.[23] Therefore, metabolites that most contribute to the separation between control and dosed samples are reduced to the horizontal or t[1] axis. Clear separation was observed between the control and dosed samples in the score plot for each herbicide (Figures 4.5B-G). Because of the use of the orthogonal component instead of another predictive component, segregation among worms as a function of dose or lack thereof is not well assessed by this approach; however, the lack of dose-dependent clustering yet significant separation observed (Figure 4.5B-G) between control and dosed-group may suggest that the MOA among doses could

be different. Nonetheless, changes in metabolite levels with respect to dose were observed in the univariate analysis (Figures 4.6 – 4.11).

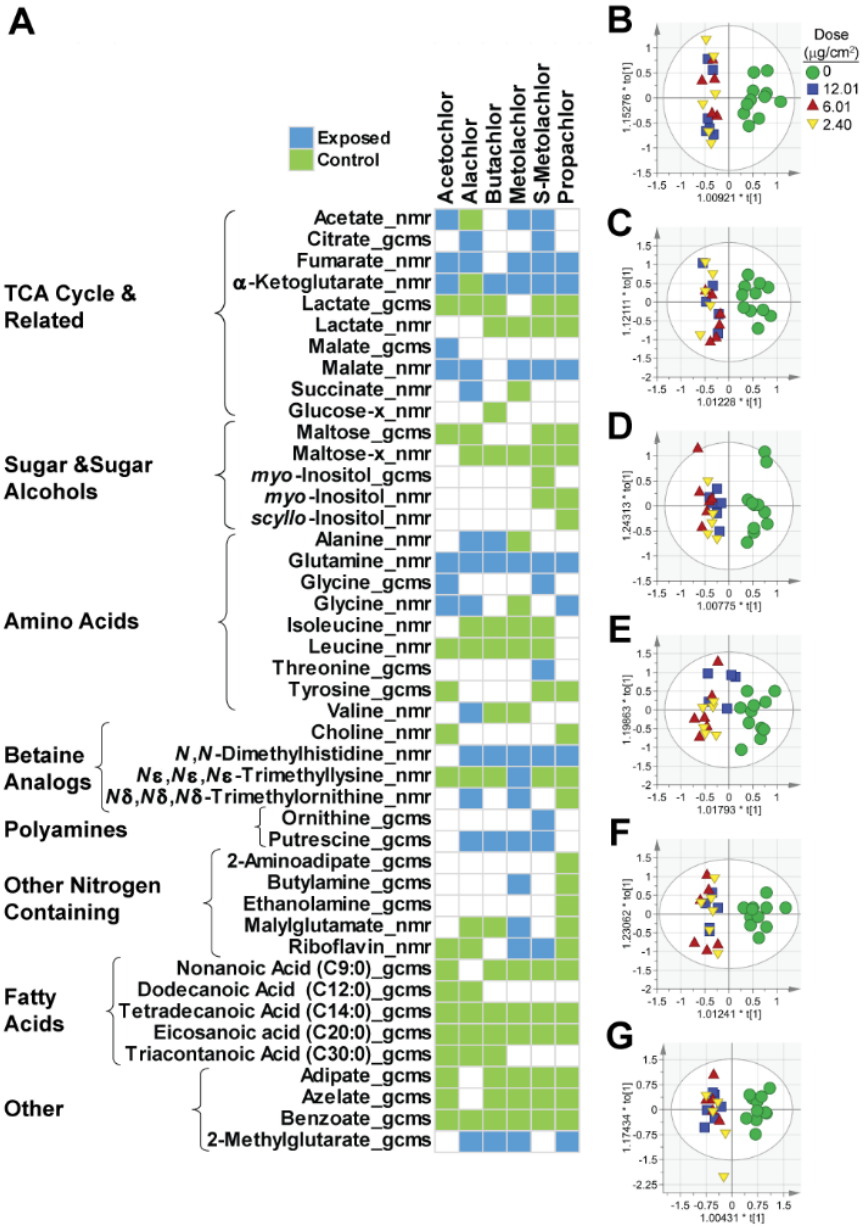


Figure 4.5. MB-OPLS-DA loadings heatmap (A) and score plots (B-G) describing the separation of CF samples from control earthworms and those exposed to chloroacetanilide herbicides. A) Heatmap of the MB-OPLS-DA loadings data depicting metabolites most important to group separation (i.e., |p(corr)| > 0.4 and VIP > 1) for controls (green), or all dosed samples (blue) for each herbicide. The MB-OPLS-DA score plots for each herbicide are plotted to the right of the heat map with data points coded as control (green), and doses at 12.01 μg/cm^2 (blue), 6.01 μg/cm^2 (red), and 2.40 μg/cm^2 (yellow) of the reported acetochlor filter test LC_{50} of 24.02 μg/cm^2.[20] The herbicide response depicted by each score plot with its corresponding CV-ANOVA results ($p \leq 0.05$, n = 6): B) acetochlor (p = 0.028); C) alachlor (p = 0.011); D) butachlor (p = 0.003); E) racemic metolachlor (p = 0.017); F) S-metolachlor (p = 0.019); and G) propachlor (p = 0.001).

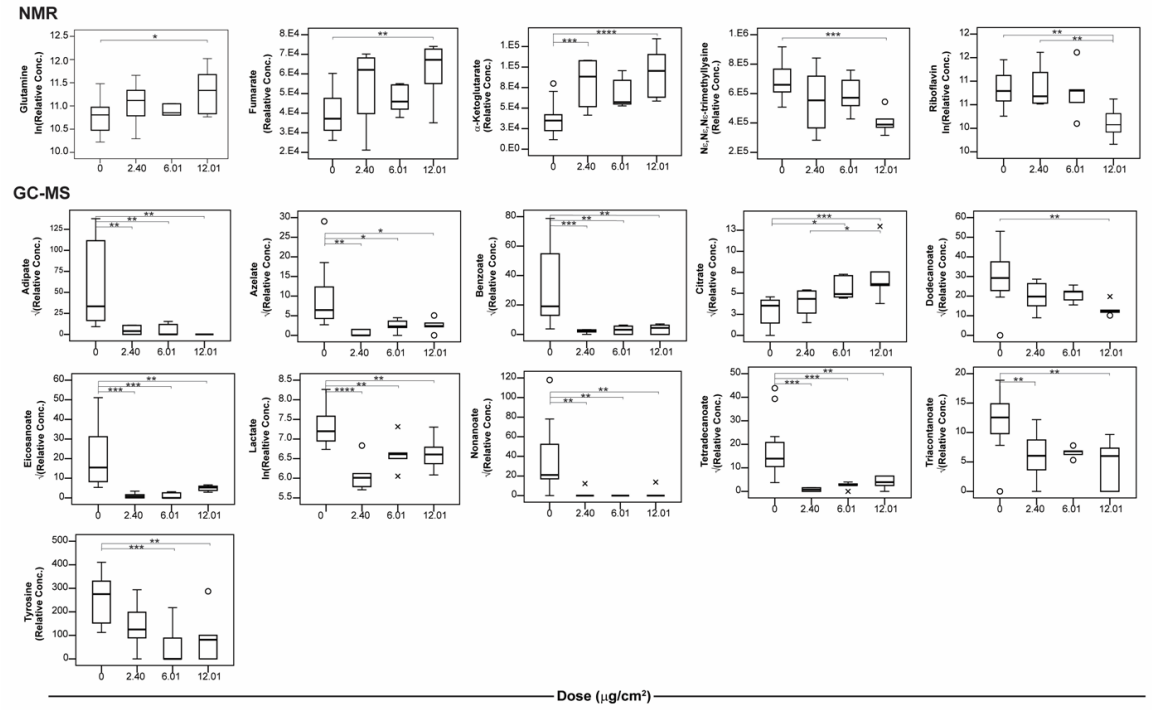


Figure 4.6. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to acetochlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = $p \leq 0.001$).

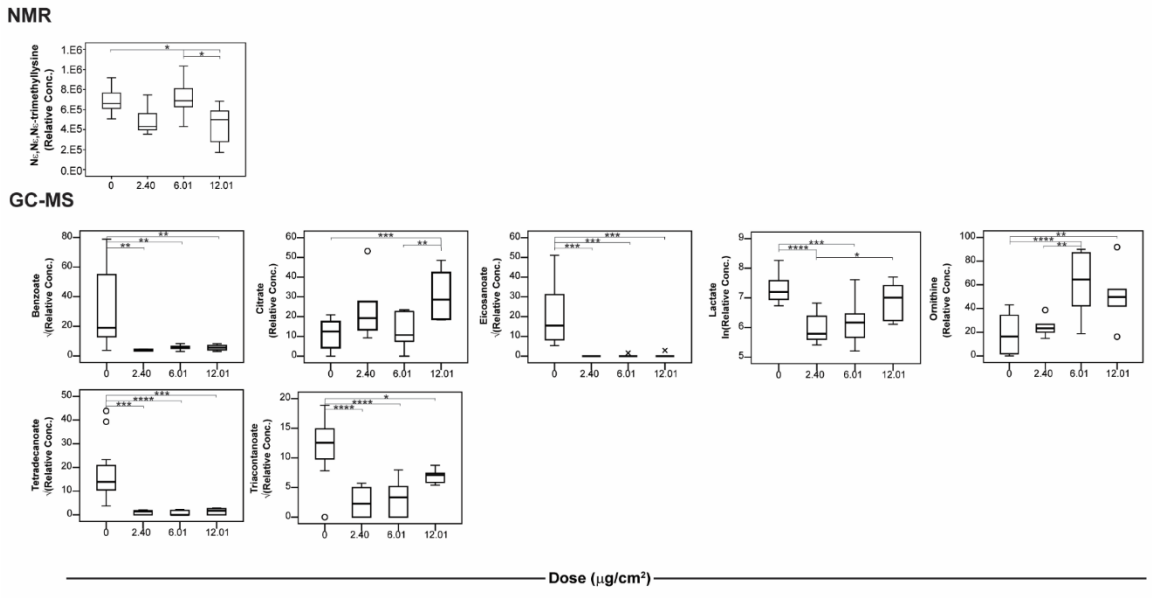


Figure 4.7. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to alachlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = $p \leq 0.001$).

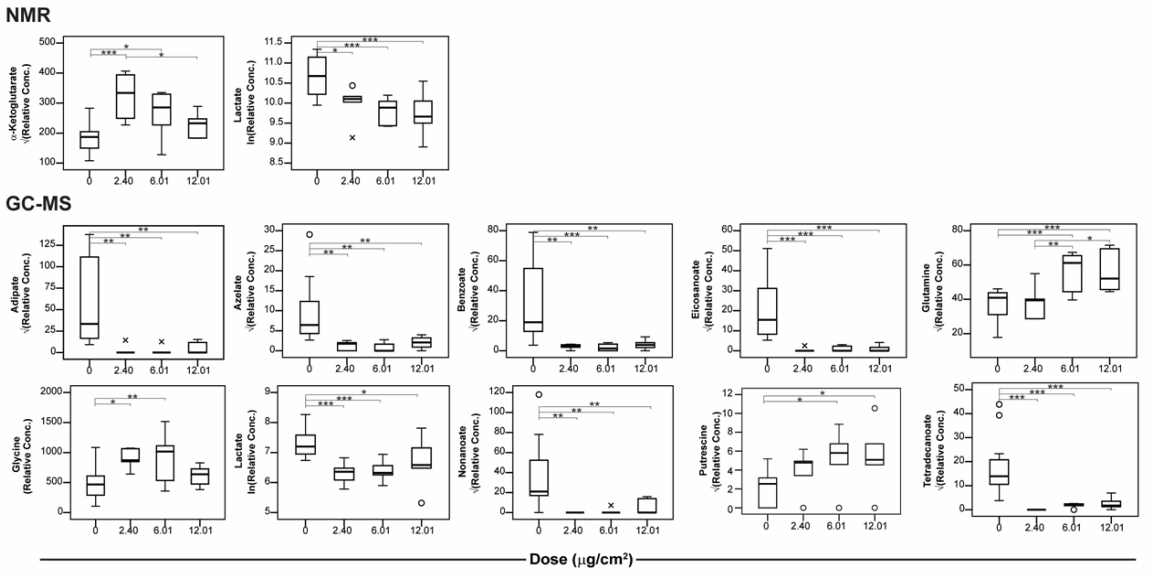


Figure 4.8. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to butachlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = $p \leq 0.001$).

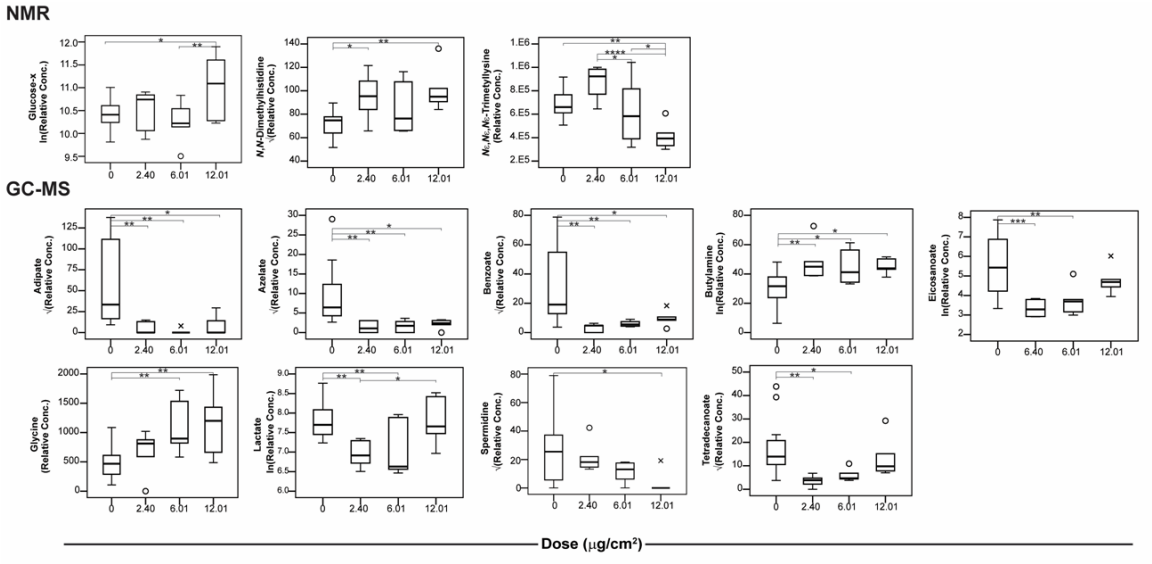


Figure 4.9. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to metolachlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = $p \leq 0.001$).

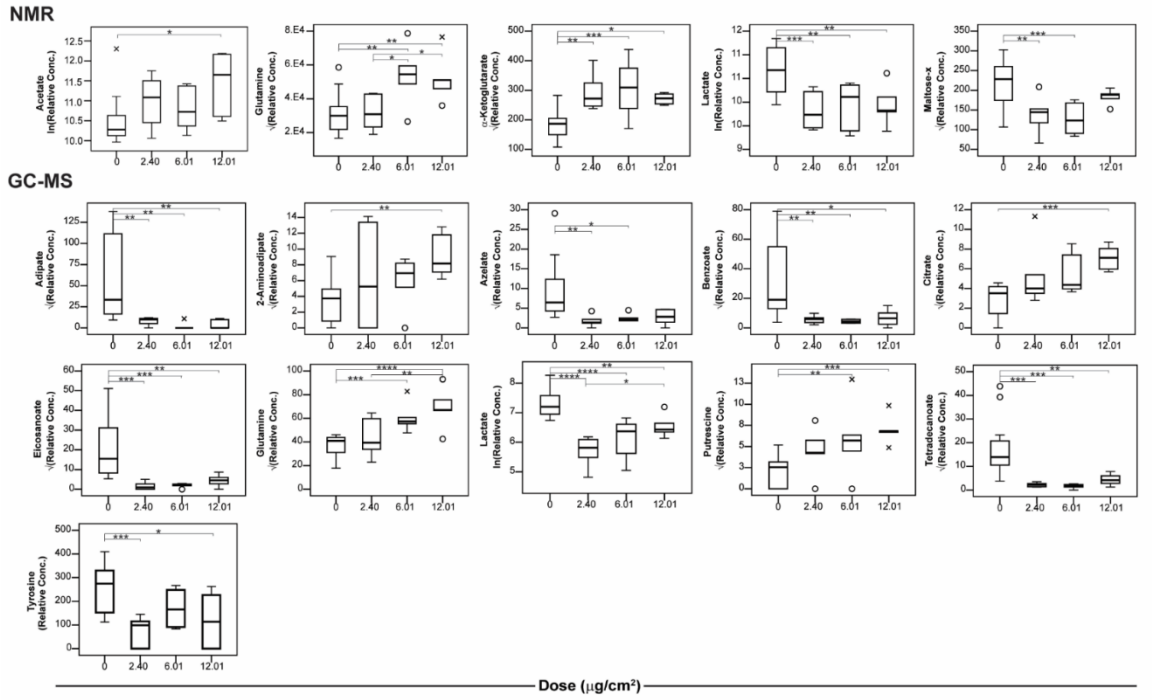


Figure 4.10. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to S-metolachlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = p 0.001).

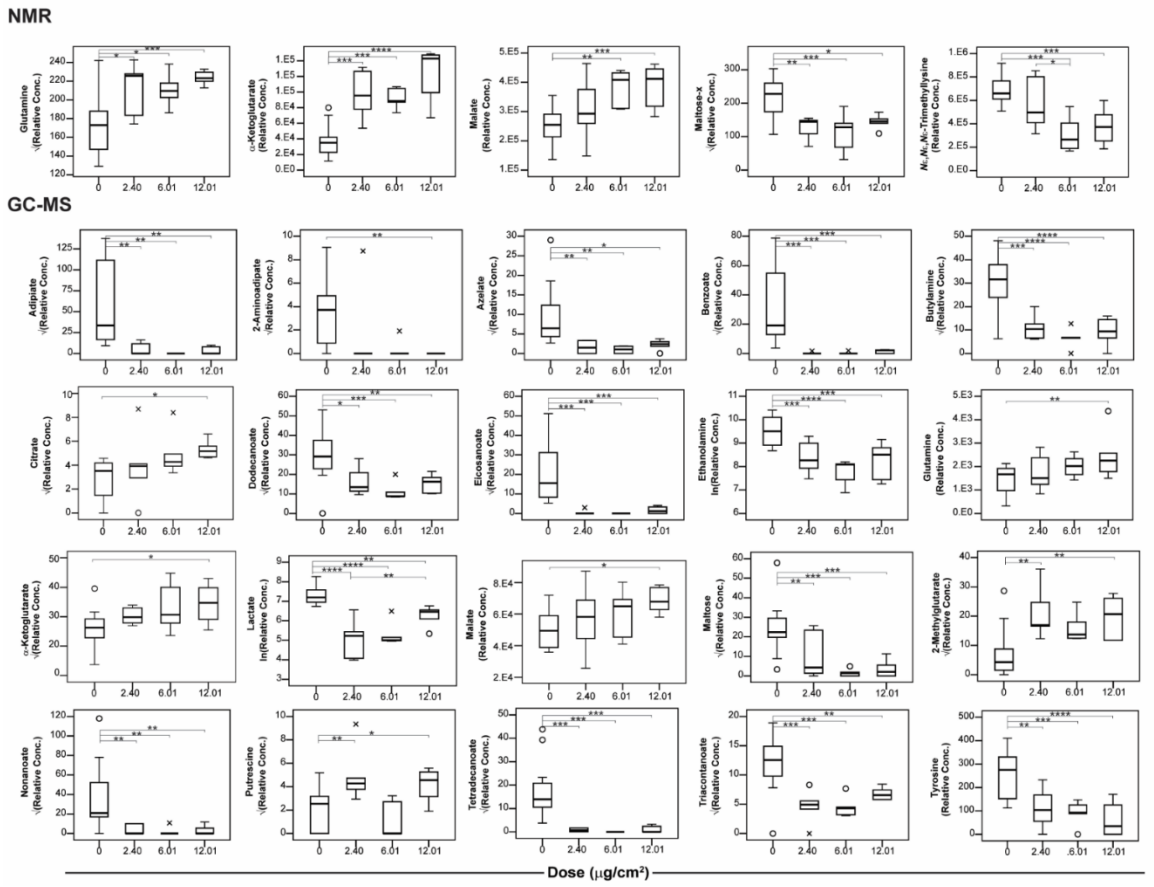


Figure 4.11. Subset of metabolic changes observed in coelomic fluid of earthworms exposed to propachlor. Tukey's HSD pairwise comparisons are annotated to show the significance level (* = $p \leq 0.10$; ** = $p \leq 0.05$; *** = $p \leq 0.01$; and **** = $p \leq 0.001$).

Univariate analysis was also performed to further interpret metabolic differences using ANOVA and Tukey's HSD ($p \leq 0.10$) with effect size calculated using Glass' Δ (Tables 4.2 – 4.7). Glass' Δ was selected because it considers variability (i.e., standard deviation) of the control group, thus a large variability in the control group would result in a smaller effect size.[25] All effect sizes calculated were $|\Delta| > 0.8$ (Tables 4.2 – 4.7), which are considered very strong.[35] A subset of metabolic changes are described in box plots (Figures 4.6 – 4.11) showing the effects of dose on metabolite levels. Univariate analysis, as summarized in the Figure 4.12 heatmap, supported many of the conclusions drawn from multivariate analysis; however, there are some differences observed between the univariate and multivariate results. Univariate analysis tests reflect independent changes in metabolite levels, meaning grouping or biochemical pathway information can be lost.[36] On the other hand, multivariate analysis can be subject to overfitting. Here, the two types of statistical analyses are not intended to validate one another as has been done in prior studies, but instead they are employed in a complementary manner to interpret the metabolic impacts of chloroacetanilide exposure in earthworms.[36]

The broadened coverage of metabolites gained by ^{1}H NMR and GC-MS analyses enhances the multivariate models and strengthens our understanding of metabolic perturbations. Differences between the results obtained by NMR and GC-MS can arise from a variety of sources including differences in sample preparation protocols, sensitivity, limits of detection, reproducibility, or data processing methods (e.g., deconvolution, normalization). For instance, lactate significantly decreases in many treatment groups (Figures 4.6 – 4.12) but is not always observed across the data produced by both

instruments due to the higher limits of quantification in the ^{1}H NMR spectra. Resonance or peak overlap is a common reason for variance between datasets, and may explain differences in glycine which is found in the crowded sugar region of ^{1}H NMR spectra.[18] Despite these limitations, we found that the complementary information gained by metabolic profiling with these orthogonal methods helped identify metabolic predictors of chloroacetanilide herbicide exposure.

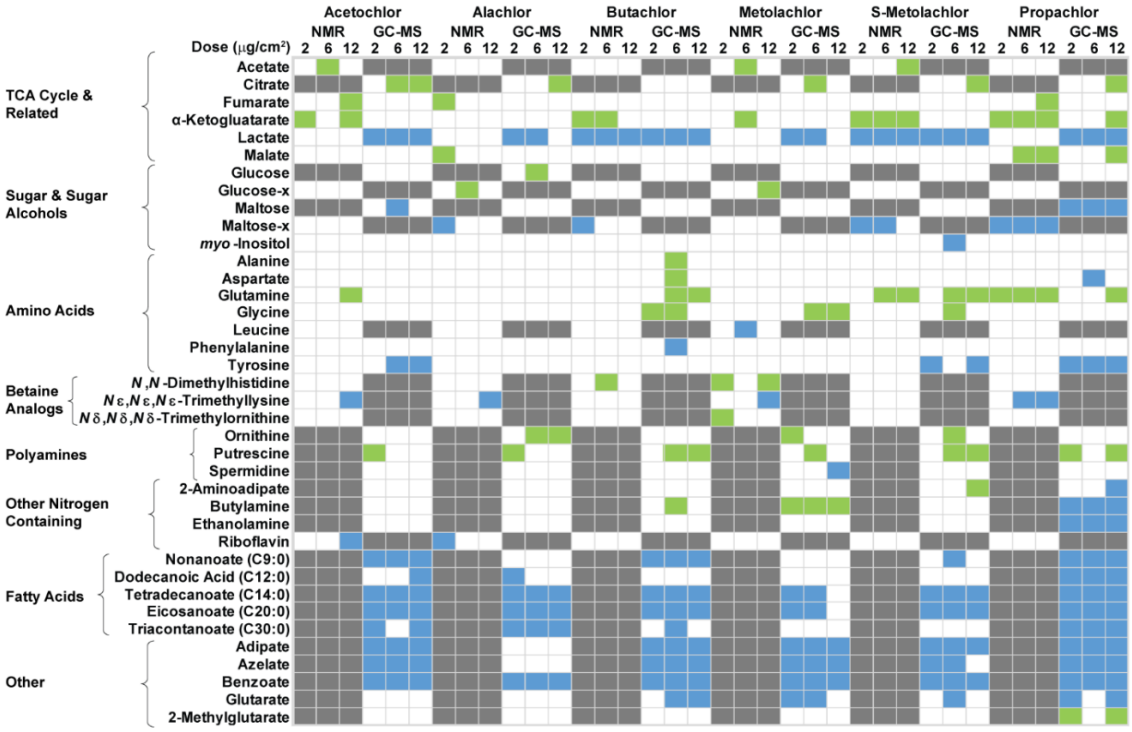


Figure 4.12. Heatmap summarizing the ANOVA and Tukey's HSD results ($p \leq 0.10$) of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to chloroacetanilide herbicides. Positive (green) or negative (blue) effect sizes are relative to the average values obtained for the controls. Very strong (i.e., $|\Delta| > 0.8$) effect sizes were obtained for all statistically significant changes (Tables 4.6 – 4.11). Metabolites not detected or quantifiable (grey) or not statistically different (white) are indicated.

Table 4.2. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to acetochlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*
Acetate			2.51	0.00					nd			
Citrate			nd						1.56	0.07	2.35	0.00
Fumarate					1.86	0.03						
α-Ketogluatarate	2.01	0.01			2.59	0.00						
Lactate							-2.48	0.00	-1.39	0.04	-1.39	0.03
Malate												
Glucose			nq									
Glucose-x									n/a			
Maltose			nq						-1.54	0.04		
Maltose-x									n/a			
myo-Inositol												
Alanine												
Aspartate												
Glutamine					1.36	0.06						
Glycine												
Leucine									nd			
Phenylalanine												
Tyrosine									-1.96	0.01	-1.65	0.02
N,*N*-Dimethylhistidine									nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine					-2.24	0.00			nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine									nd			
Ornithine			nd									
Putrescine			nq				3.58	0.03				
Spermidine			nq									
2-Aminoadipate			nd									
Butylamine			nd									
Ethanolamine			nd									
Riboflavin					-1.82	0.01			nd			
Nonanoate (C9:0)			nd				-1.03	0.02	-1.09	0.03	-1.02	0.03
Dodecanoate (C12:0)			nd								-1.23	0.01
Tetradecanoate (C14:0)			nd				-1.37	0.00	-1.23	0.01	-1.11	0.01
Eicosanoate (C20:0)			nd				-1.25	0.01	-1.25	0.01	-1.01	0.03
Triacontanoate (C30:0)			nd				-1.19	0.05			-1.44	0.01
Adipate			nd				-1.04	0.02	-1.03	0.04	-1.14	0.01
Azelate			nd				-1.14	0.01	-0.89	0.09	-0.88	0.07
Benzoate			nd				-1.16	0.01	-1.12	0.02	-1.10	0.02
Glutarate			nd									
2-Methylglutarate			nd									

*nd = not detected; nq = not quantifiable; n/a = not applicable

Table 4.3. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to alachlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*		Δ	*p*	Δ	*p*	Δ
Acetate									nd			
Citrate			nd								2.58	0.01
Fumarate	2.01	0.03										
α-Ketogluatarate												
Lactate							-2.71	0.00	-2.22	0.01		
Malate	1.97	0.02										
Glucose			nq						2.11	0.03		
Glucose-x			1.63	0.02					n/a			
Maltose			nq									
Maltose-x	-1.78	0.03							n/a			
myo-Inositol												
Alanine												
Aspartate												
Glutamine												
Glycine												
Leucine									nd			
Phenylalanine												
Tyrosine												
N,*N*-Dimethylhistidine									nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine					-1.94	0.05			nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine									nd			
Ornithine			nd						2.50	0.00	1.88	0.02
Putrescine			nq				3.50	0.01				
Spermidine			nq									
2-Aminoadipate			nd									
Butylamine			nd									
Ethanolamine			nd									
Riboflavin	-2.09	0.02							nd			
Nonanoate (C9:0)			nd									
Dodecanoate (C12:0)			nd				-0.99	0.05				
Tetradecanoate (C14:0)			nd				-1.34	0.00	-1.37	0.00	-1.31	0.00
Eicosanoate (C20:0)			nd				-1.32	0.00	-1.31	0.00	-1.29	0.00
Triacontanoate (C30:0)			nd				-1.91	0.00	-1.76	0.00	-1.02	0.06
Adipate			nd									
Azelate			nd									
Benzoate			nd				-1.09	0.01	-1.02	0.02	-1.03	0.02
Glutarate			nd									
2-Methylglutarate			nd									

*nd = not detected; nq = not quantifiable; n/a = not applicable

Table 4.4. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to butachlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*
Acetate									nd			
Citrate			nd									
Fumarate												
α-Ketogluatarate	2.44	0.00	1.43	0.09								
Lactate	-1.36	0.06	-1.68	0.01	-1.85	0.00	-2.03	0.01	-1.87	0.01	-1.33	0.09
Malate												
Glucose			nq									
Glucose-x									n/a			
Maltose			nq									
Maltose-x	-2.53	0.01							n/a			
myo-Inositol												
Alanine									1.74	0.07		
Aspartate									2.58	0.00		
Glutamine									2.12	0.01	2.05	0.01
Glycine							1.37	0.07	1.46	0.03		
Leucine									nd			
Phenylalanine									-1.28	0.04		
Tyrosine												
N,*N*-Dimethylhistidine			1.86	0.10					nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine									nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine									nd			
Ornithine			nd									
Putrescine			nq						1.80	0.07	1.82	0.06
Spermidine			nq									
2-Aminoadipate			nd									
Butylamine			nd						1.32	0.09		
Ethanolamine			nd									
Riboflavin									nd			
Nonanoate (C9:0)			nd				-1.09	0.03	-1.06	0.02	-0.94	0.04
Dodecanoate (C12:0)			nd									
Tetradecanoate (C14:0)			nd				-1.35	0.00	-1.12	0.00	-0.94	0.01
Eicosanoate (C20:0)			nd				-1.29	0.01	-1.27	0.00	-1.26	0.01
Triacontanoate (C30:0)			nd						-1.23	0.05		
Adipate			nd				-1.08	0.03	-1.10	0.02	-1.05	0.02
Azelate			nd				-1.04	0.04	-1.11	0.01	-0.95	0.05
Benzoate			nd				-1.13	0.02	-1.15	0.01	-1.08	0.02
Glutarate			nd									
2-Methylglutarate			nd									

*nd = not detected; nq = not quantifiable; n/a = not applicable

Table 4.5. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to metolachlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*		Δ	*p*	Δ	*p*	Δ
Acetate			1.56	0.06					nd			
Citrate			nd						1.51	0.04		
Fumarate												
α-Ketogluatarate			1.47	0.09								
Lactate							-1.74	0.02	-1.60	0.03		
Malate												
Glucose			nq									
Glucose-x					1.80	0.09			n/a			
Maltose			nq									
Maltose-x									n/a			
myo-Inositol												
Alanine												
Aspartate												
Glutamine												
Glycine									1.98	0.04	2.25	0.03
Leucine			-1.13	0.06					nd			
Phenylalanine												
Tyrosine												
N,*N*-Dimethylhistidine	1.96	0.06			2.40	0.01			nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine					-2.20	0.02			nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine	1.73	0.02							nd			
Ornithine			nd				1.48	0.08				
Putrescine			nq						2.65	0.00		
Spermidine			nq								-0.92	0.10
2-Aminoadipate			nd									
Butylamine			nd				1.49	0.02	1.20	0.08	1.26	0.08
Ethanolamine			nd									
Riboflavin									nd			
Nonanoate (C9:0)			nd									
Dodecanoate (C12:0)			nd									
Tetradecanoate (C14:0)			nd				-1.14	0.02	-0.96	0.07		
Eicosanoate (C20:0)			nd				-1.41	0.00	-1.15	0.02		
Triacontanoate (C30:0)			nd									
Adipate			nd				-1.05	0.02	-1.11	0.01	-0.97	0.06
Azelate			nd				-1.03	0.03	-1.00	0.03	-0.93	0.07
Benzoate			nd				-1.11	0.01	-1.01	0.03	-0.87	0.10
Glutarate			nd				-0.81	0.09	-0.81	0.09		
2-Methylglutarate			nd									

*nd = not detected; nq = not quantifiable; n/a = not applicable

Table 4.6. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to S-metolachlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*
Acetate					1.31	0.09			nd			
Citrate			nd								2.37	0.01
Fumarate												
α-Ketogluatarate	1.91	0.02	2.17	0.01	1.53	0.09						
Lactate	-1.67	0.01	-1.42	0.03	-1.41	0.04	-3.23	0.00	-2.36	0.00	-1.54	0.05
Malate												
Glucose			nq									
Glucose-x									n/a			
Maltose			nq									
Maltose-x	-1.23	0.03	-1.41	0.01					n/a			
myo-Inositol									-1.09	0.10		
Alanine												
Aspartate												
Glutamine			1.55	0.02	1.49	0.04			2.52	0.01	3.46	0.00
Glycine									2.05	0.01		
Leucine									nd			
Phenylalanine												
Tyrosine							-1.80	0.00			-1.36	0.06
N,*N*-Dimethylhistidine									nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine									nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine									nd			
Ornithine			nd						1.53	0.03		
Putrescine			nq						2.10	0.03	2.71	0.01
Spermidine			nq									
2-Aminoadipate			nd								2.02	0.05
Butylamine			nd									
Ethanolamine			nd									
Riboflavin									nd			
Nonanoate (C9:0)			nd						-0.98	0.03		
Dodecanoate (C12:0)			nd									
Tetradecanoate (C14:0)			nd				-1.25	0.01	-1.30	0.00	-1.07	0.03
Eicosanoate (C20:0)			nd				-1.22	0.01	-1.19	0.01	-1.04	0.03
Triacontanoate (C30:0)			nd									
Adipate			nd				-0.98	0.03	-1.10	0.01	-1.06	0.03
Azelate			nd				-0.99	0.04	-0.89	0.07		
Benzoate			nd				-1.02	0.03	-1.07	0.02	-0.98	0.05
Glutarate			nd						-0.81	0.10		
2-Methylglutarate			nd									

*nd = not detected; nq = not quantifiable; n/a = not applicable

Table 4.7. ANOVA and Tukey's HSD results ($p \leq 0.10$) and Glass' Δ of metabolic perturbations detected by ^{1}H NMR and GC-MS for earthworms exposed to propachlor.

	NMR						GC-MS					
Dose (μg/cm^2)	2.40		6.01		12.01		2.40		6.01		12.01	
	Δ	*p*	Δ	*p*	Δ	*p*		Δ	*p*	Δ	*p*	Δ
Acetate									nd			
Citrate			nd								1.37	0.08
Fumarate					1.29	0.09						
α-Ketogluatarate	2.74	0.00	2.52	0.00	4.02	0.00					1.43	0.08
Lactate							-4.47	0.00	-4.01	0.00	-2.09	0.02
Malate			2.05	0.03	2.19	0.01					1.49	0.07
Glucose			nq									
Glucose-x									n/a			
Maltose			nq				-1.09	0.05	-1.70	0.00	-1.56	0.00
Maltose-x	-1.44	0.02	-1.66	0.01	-1.14	0.05			n/a			
myo-Inositol												
Alanine												
Aspartate									-1.28	0.05		
Glutamine	1.04	0.09	1.05	0.09	1.40	0.01					1.67	0.04
Glycine												
Leucine									nd			
Phenylalanine												
Tyrosine							-1.46	0.01	-1.66	0.01	-1.96	0.00
N,*N*-Dimethylhistidine									nd			
*N*ε,*N*ε,*N*ε-Trimethyllysine			-2.95	0.00	-2.47	0.01			nd			
*N*δ,*N*δ,*N*δ-Trimethylornithine									nd			
Ornithine			nd									
Putrescine			nq				1.57	0.02			1.19	0.10
Spermidine			nq									
2-Aminoadipate			nd								-1.25	0.04
Butylamine			nd				-1.55	0.00	-1.91	0.00	-1.68	0.00
Ethanolamine			nd				-1.88	0.01	-2.90	0.00	-2.03	0.00
Riboflavin									nd			
Nonanoate (C9:0)			nd				-0.99	0.03	-1.03	0.04	-1.01	0.03
Dodecanoate (C12:0)			nd				-1.00	0.05	-1.37	0.01	-1.05	0.04
Tetradecanoate (C14:0)			nd				-1.37	0.00	-1.43	0.00	-1.35	0.00
Eicosanoate (C20:0)			nd				-1.29	0.00	-1.32	0.01	-1.22	0.01
Triacontanoate (C30:0)			nd				-1.48	0.00	-1.50	0.00	-1.05	0.04
Adipate			nd				-1.05	0.02	-1.14	0.02	-1.08	0.02
Azelate			nd				-1.00	0.03	-1.08	0.03	-0.92	0.05
Benzoate			nd				-1.22	0.01	-1.22	0.01	-1.18	0.01
Glutarate			nd				-0.81	0.09			-0.81	0.09
2-Methylglutarate			nd				1.55	0.01			1.45	0.02

*nd = not detected; nq = not quantifiable; n/a = not applicable

4.3.1. Measured Biochemical Perturbations

Metabolic profiling revealed that chloroacetanilide herbicides induced perturbations in lipid, energy, and nitrogen metabolism in the *E. fetida* CF. Levels of eicosanoate (C20:0) significantly decrease (Figure 4.12) in exposed worms in all herbicide groups and triacontanoate (C30:0) significantly decreases with very strong effect sizes in worms exposed to acetochlor, alachlor, and propachlor (Figures 4.6 – 4.11, Tables 4.2 – 4.7). MB-OPLS-DA (Figure 4.5A) shows that fatty acids are drivers of group separation, which is demonstrated further in Figures 4.6 – 4.11 and Tables 4.2 – 4.7 as levels of the shorter-chain fatty acids nonanoate (C9:0), dodoecanoate (C12:0), and tetradecanoate (C14:0) decrease in the dosed groups for all herbicides. Decreased fatty acid levels suggest an increase in fatty acid β-oxidation to support an increased energy demand in exposed worms, as represented in biochemical map shown in Figure 4.13.

β-oxidation requires carnitine to facilitate transport of fatty acids into the mitochondria, where they are oxidized to acetyl-CoA and used to generate adenosine triphosphate (ATP) via the tricarboxylic (TCA) cycle and electron transport chain (Figure 4.13).[37] The carnitine precursor, TML, is a control-related driver of MB-OPLS-DA separation, except in the metolachlor model, (Figure 4.5A) and significantly decreases in earthworms exposed to acetochlor, metolachlor, and propachlor (Figures 4.6, 4.8, 4.10, and 4.12). While TML is associated with the controls, glycine is a byproduct of carnitine biosynthesis (Figure 4.13) and is associated with several exposed groups in MB-OPLS-DA (Figure 4.5A).[37] In the univariate analysis, glycine only increases significantly in the GC-MS butachlor and metolachlor samples (Figure 4.6). The NMR-detected TCA intermediate

α-ketoglutarate results contributed to all group separations except alachlor (Figure 4.5A) and a significant increase in α-ketoglutarate levels was observed in some groups (Figures 4.6 – 4.12, Tables 4.2 – 4.7). α-Ketoglutarate is converted to succinate in two steps of carnitine biosynthesis (Figure 4.13). Succinate contributed to separation in the alachlor, metolachlor, and propachlor models (Figure 4.5A), but changes were not significant as evaluated by univariate analysis. The significance of TML, glycine, and α-ketoglutarate within the multivariate and univariate analyses suggests that carnitine biosynthesis is upregulated to drive β-oxidation of fatty acids to fuel the TCA cycle.

Further reinforcing an increase in energy demand induced by chloroacetanilide exposure, maltose (GC-MS) and maltose-x (i.e., overlapped NMR resonances of maltose, maltotriose, and other sugars (Figure 4.2)) decrease in S-metolachlor and propachlor exposed worms (Figures 4.10 – 4.12, Table 4.6 – 4.7). MB-OPLS-DA (Figure 4.5A) also shows that maltose-x is important for separation of controls in all models except acetochlor exposed worms, while maltose is important in the acetochlor, alachlor, S-metolachlor, and propachlor models. Decreases in lactate, the reduced form of pyruvate, is seen consistently in exposure of all herbicides, suggesting that lactate is oxidized to pyruvate then acetyl-CoA and pushed into the TCA cycle.[38] Activation of acetate is another acetyl-CoA source and acetate is important to separation in some models (Figure 4.5A). TCA cycle intermediates citrate, α-ketoglutarate, succinate, fumarate, and malate (Figure 4.13) are detected in CF samples and all were flagged important by multivariate analysis. Significant increases in citrate levels are observed in the acetochlor, alachlor, metolachlor, and S-metolachlor exposures (Figures 4.6, 4.7, and 4.9 – 4.12, Table 4.2 – 4.7) and contribute to

separation of dosed and control samples in the acetochlor and S-metolachlor models (Figure 4.5A), further supporting the induced metabolic pressure on the earthworm's energy demand by chloroacetanilide herbicides.

Alterations in nitrogen metabolism in the CF of exposed earthworms were also observed. Sequestration of free ammonium from increased catabolism could be responsible for the increased levels of glutamine observed in all models (Figures 4.5A and 4.12).[38] In worms exposed to S-metolachlor, significant decreases and strong effect sizes were observed for tyrosine (Figures 4.10 and 4.12, Table 4.6) in worms exposed to propachlor at all doses, acetochlor at the high and middle dose, and S-metolachlor at the lowest dose. The betaine (i.e., trimethylamine) analogs DMH and TMO were also important to separation in several models (Figure 5=4.5A). TML's role is not well understood but it is hypothesized to play an osmoprotective role to help earthworms tolerate extremes in soil moisture content. As TML is also critical for carnitine biosynthesis,[4] we postulate that decreased levels are associated with increased lipid β-oxidation.[18, 39] This hypothesis is further supported by the increase in putrescine and ornithine levels (Figures 4.6 – 4.12, Tables 4.2 – 547) and their association with the exposed groups (Figure 4.5A), given that increased polyamines serve a protective role against metabolic pressures like osmotic, oxidative, and thermal stress.[40] Although changes in nitrogen metabolism are observed in chloroacetanilide herbicide exposed earthworms, a more targeted approach would need to be employed to ascertain the exact biochemical pathways or processes impacted.

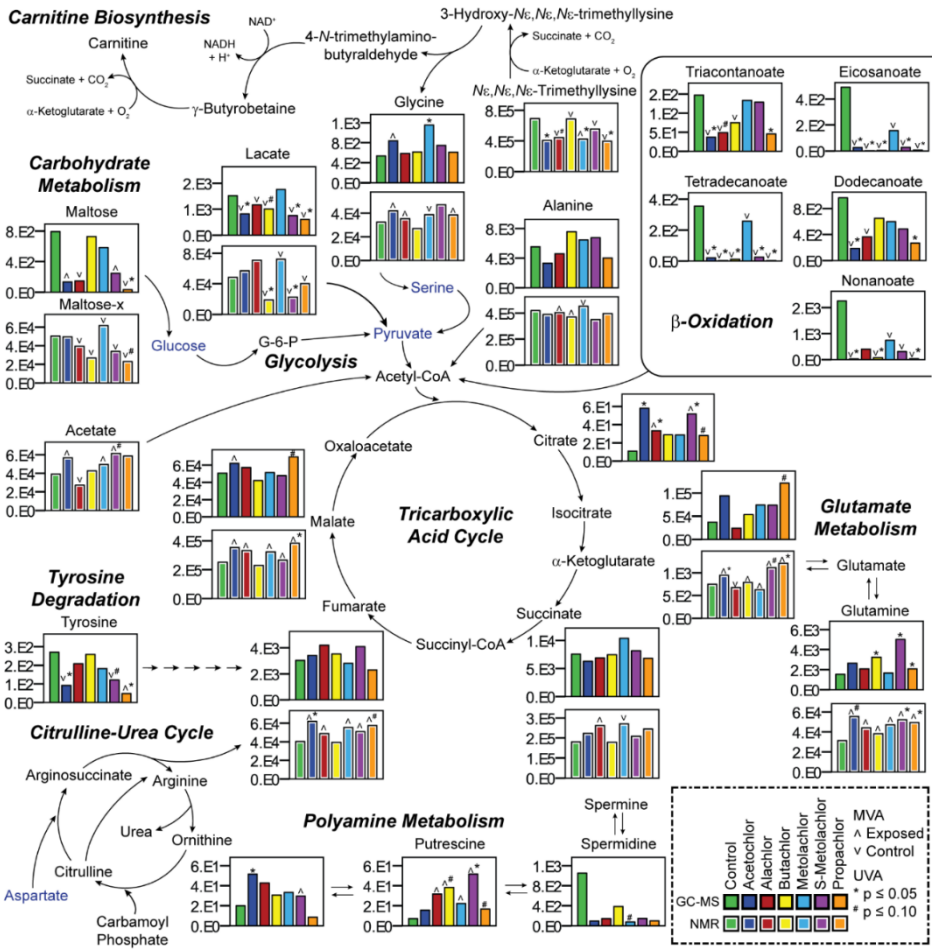


Figure 4.13. Diagram of the biochemical pathways perturbed in earthworms by chloroacetanilide herbicides and bar graphs, reported in relative concentration, for the high dose of each treatment group. Treatment is indicated by color: control (green); acetochlor (dark blue); alachlor (red); butachlor (yellow); racemic metolachlor (light blue); S-metolachlor (purple); and propachlor (orange). Bars representing GC-MS results are outlined in black, while NMR results are indicated by a white outline. Metabolites that are were identified as most important to segregation of exposed (∧) or controls (∨) in MB-OPLS-DA are marked, in addition to metabolites that differ from the controls as $p \leq 0.10$ (#) and $p \leq 0.05$ (*), which is further described in the box plots in Figures 4.6 – 4.11.

4.3.2. Herbicide Mode of Action

Chloroacetanilides are preemergent herbicides designed to inhibit early plant growth by disrupting the synthesis and elongation of very-long-chain fatty acids.[5] Elongation extends C16:0 and C18:0 fatty acids to larger chains. As observed in Figure 4.13, levels of fatty acids decrease in earthworms exposed to chloroacetanilide herbicides. Corresponding decreases in TML levels could arise from an increased demand for carnitine to facilitate the transport of fatty acids into the mitochondria to undergo β-oxidation to support an increased energy demand as a result of exposure. Consistent with these findings, increased transcription of genes related to lipid metabolism and transport was observed in alachlor exposed yeast (*Saccharomyces cerevisiae*) with upregulation in the *CRC1* gene suggesting that alachlor induced interference in acylcarnitine ester transfer resulting in increased fatty acid β-oxidation.[41] Moreover, chemoproteomic profiling of the livers of mice exposed to acetochlor revealed an inhibition of fatty acid oxidation and as a result increased levels of free fatty acids and other lipid species.[42] Because our observations are closer to those reported for *S. cerevisiae*, we postulate that the observations in the mouse study may be attributed to phylogenetic or toxicity differences. Nonetheless, our study supports the evidence that chloroacetanilide exposure disrupts lipid metabolism, impacting pathways in nontarget organisms in a manner similar to the mechanism of herbicidal activity.

The metabolic pressure exerted by chloroacetanilide herbicides on earthworm energy demand is further corroborated in the literature. Proteomic and metabolomic analysis of the fungus *Paecilomyces marquandii* exposed to alachlor revealed upregulation

in TCA cycle intermediates and proteins, which was similarly observed in transcriptomic profiling of *S. cerevisiae* exposed to alachlor.[41, 43] Additionally, significant decreases in maltose, glucose, and glucose-6-phosphate were observed in the ^{1}H NMR metabolic profiles of goldfish (*Carassius auratus*) exposed to butachlor.[38, 44]

Changes in energy demand could also affect osmolyte levels. The observed alterations of DMH and TMO levels could be attributed to the funneling of metabolites into the TCA cycle, where the positively charged DMH and TMO could help maintain charge balance in the CF compensating for the increased organic acid content. Additionally, changes in osmolyte levels could also be suggestive of oxidative stress, as observed in butachlor exposed goldfish where increased levels of glutathione, taurine, and betaine were observed.[44] This observation is consistent with the reported increase in the levels of proteins involved in combating reactive oxygen species in *P. marquandii* exposed to alachlor.[43] Changes in polyamine levels may also reflect oxidative stress, which can induce DNA damage. Acetochlor has been demonstrated to induce DNA damage in earthworm coelomocytes,[45] while butachlor induced a loss of chromatin in earthworm nuclei.[46] Taken together, these results support our observations that exposure of nontarget organisms like earthworms to chloroacetanilide herbicides alters lipid metabolism and induces oxidative stress, spiking energy demands to combat the applied stress.

4.3.3. Structure Activity Relationships

A question raised by this study was whether a relationship exists between the metabolic response and the structural differences within the herbicide class (Figure 4.1).

Similar pathways were impacted by all the chloroacetanilides on the earthworm metabolome: carbohydrate metabolism, carnitine biosynthesis, energy metabolism, and lipid metabolism. Within these pathways, however, unique molecular profiles or phenotypes were observed as the impacted metabolites affected within those pathways vary by compound. For instance, the loadings plot heatmap (Figure 4.5A) showed malylglutamate to be important for separation in the alachlor, butachlor, metolachlor, and propachlor models, but it was only associated with the exposed group in the metolachlor model. Similarly, in the S-metolachlor MB-OPLS-DA (Figure 4.5A), threonine was associated with exposed worms, which is not observed in the other treatment groups.

The idea that each herbicide may induce a specific metabolic profile is further supported in the multiblocked-partial least squares-discriminant analysis (MB-PLS-DA) score plots (Figures 4.14A-B) of all samples as a function of treatment (i.e., doses were combined). Multiblocked-partial least squares-hierarchical cluster analysis (MB-PLS-HCA) of the model (Figure 4.14C) further highlights treatment clustering, showing two major clusters. Propachlor forms its own cluster high up the dendrogram, at 560, indicating that its quite different from the other treatment groups (Figure 4.14C). The second major cluster consists of several subsequent clusters, with the first indicating that control samples cluster separately from the other treatment groups. The next sub-cluster highlights the similarity among acetochlor, butachlor, metolachlor, and S-metolachlor groups (Figure 4.14C), which are well overlapped in the score plots (Figures 4.14A-B). Overall, MB-PLS-HCA shows that the treatment groups cluster with themselves well, indicating that there is some specificity in the metabolic response of earthworm CF to the chloroacetanilides.

The MB-PLS-DA score plots in Figures 4.14A-B further illustrate strong clustering and separation of propachlor from the rest of the treatment groups. Interestingly, propachlor is structurally quite different than the other chloroacetanilide herbicides tested. It lacks methyl or ethyl substituents on the benzene ring and has an isopropyl group attached to the chloroacetamide rather than an ester (Figure 4.1), which may influence its effects within some biochemical pathways. More changes in metabolite levels that were significant (Figures 4.11 and 4.12, Table 4.7) are observed at the low and middle dose in propachlor exposed earthworms compared to the other herbicides. Additionally, more metabolites were affected by propachlor exposure. A decrease in butylamine and ethanolamine (Figures 4.11 and 4.12) was observed only in propachlor exposed worms, which was also important in MB-OPLS-DA loadings heatmap (Figure 4.5A). Figure 4.5A further highlights differences between propachlor and the other herbicides. It is the only model that associates 2-aminoadipate, butylamine, ethanolamine, and *scyllo*-inositol with control worms. Additionally, the propachlor model (Figure 4.5G) was the most significant model by CV-ANOVA with $p = 0.001$. Although increased metabolic perturbations are observed in propachlor exposed worms, it cannot necessarily be concluded that propachlor is the most toxic due to the design of the experiment.

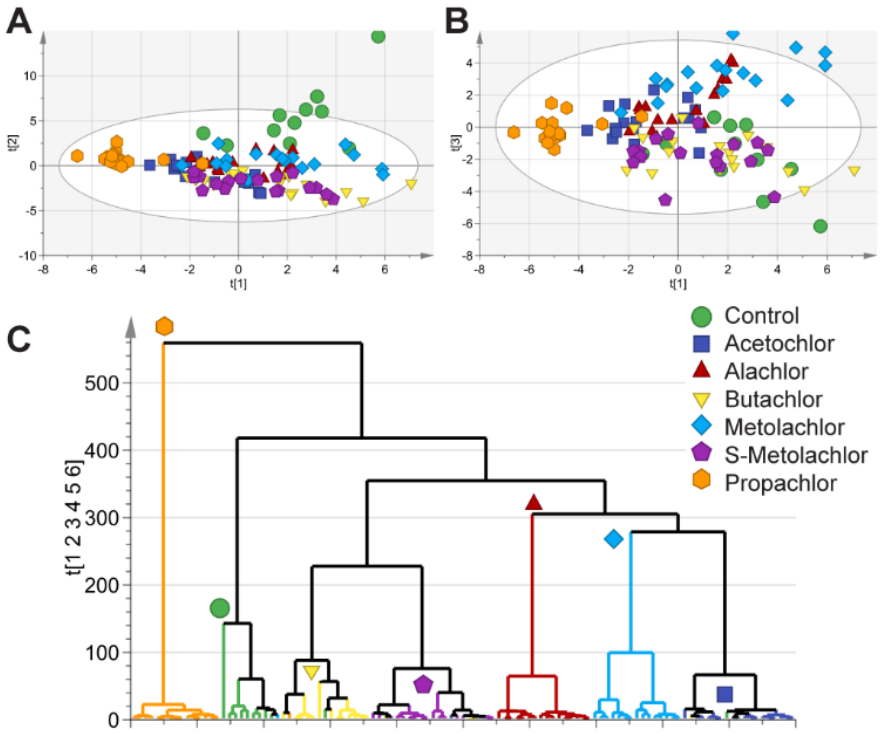


Figure 4.14. MB-PLS-DA of all data as a function of treatment. A) Score plot of t[1] and t[2]; B) score plot of t[1] and t[3]; and C) dendrogram describing the MB-PLS-HCA of treatments. Hierarchal clustering was calculated using Ward's methods and sorted by size.

4.3.4. Stereospecific Metabolic Impacts

Dual Magnum®, the metolachlor formation that contains primarily the S-enantiomer (Figure 4.1E), was introduced to the market in 1997. Because the S-enantiomer of metolachlor is the active herbicidal component, this formulation resulted in reduced amounts of material applied to achieve the same effect.[47] Lower application rates are environmentally beneficial due to chloroacetanilide groundwater mobility and, in theory, less material introduced into the environment should reduce risks to nontarget organisms. In contrast, however, *E. fetida* was found to be more sensitive to racemic metolachlor than

S-metolachlor in avoidance tests, and racemic metolachlor also induced greater changes in bodyweight and higher catalase and cellulase activity, suggesting that racemic metolachlor was more toxic than the S-enantiomer.[48] Our results, shown in Figure 4.14, reveal that racemic and S-metolachlor cluster separately in the MB-PLS-HCA (Figure 4.14C), which motivated us to examine the stereospecific metabolic impacts between the two datasets. Both racemic and S-metolachlor give rise to consistent results for acetate, fumarate, α-ketoglutarate, lactate, malate, and maltose-x (Figure 4.5A); however, citrate is important to the exposed group only in the S-metolachlor model, and succinate is important to the control group only in the racemic metolachlor model. *Myo*-inositol contributes to separation of the controls in the S-metolachlor model. A notable difference in the multivariate and univariate analysis (Figures 4.5, 4.9, 4.10, and 4.12, Tables 4.5 – 4.6) is the elevation of glutamine in S-metolachlor exposed worms but not in the racemic metolachlor exposed groups. Additionally, threonine, tyrosine, and ornithine contribute to the separation of the control group from S-metolachlor exposed worms, while alanine, butylamine, malylglutamate, 2-methylglutarate, TMO, and valine are critical to the separation of racemic metolachlor exposed worms and the controls. Though we did not explicitly compare toxicity in our experiments in terms of lethality, we can conclude that racemic metolachlor and S-metolachlor affect the earthworm metabolome differently. These findings suggest that metabolomic perturbations can distinguish differences in environmental impacts, not only in structurally analogous compounds, but potentially differences in enantiomeric ratios as well, an idea that warrants further investigation.

4.4. Conclusions

Expanding the knowledge of how the metabolome responds to chemical exposures is propelling the use of metabolomics in understanding toxicity, describing adverse outcome pathways, and monitoring environmental insults. Earthworms are the choice candidate to signal the health of terrestrial ecosystems and earthworm metabolomics is demonstrated to be a useful tool to evaluate chemical stress.[2-3] Herein, CF is further supported as a beneficial metabolite pool to evaluate toxicity and monitor earthworm health. Still, studies probing metabolic perturbations in earthworm CF are limited, though it offers a noninvasive, nonlethal, and rich metabolite source with greater temporal resolution than whole worm analyses.

Investigation into the metabolic impacts of chloroacetanilide herbicides on the earthworm CF revealed metabolic signatures that suggest that their MOA is not specific to plants. The herbicides disrupt fatty acid metabolism and the elongation or synthesis of very-long fatty acids. Carnitine biosynthesis and fatty acid β-oxidation increased in earthworms exposed to chloroacetanilides. In addition, exposed earthworms showed increased levels of TCA cycle intermediates and decreased sugar and lactate levels suggesting an increased energy demand. Corroborating results have been observed in multiple species (*S. cerevisiae*, *P. marquandii*, goldfish, and plants) and biological platforms (e.g., metabolomics, proteomics, and transcriptomics).[41, 43-44, 49] Even in the mouse liver where fatty acids were not oxidized, an increase in lipid species and impaired fatty acid metabolism was observed,[42] demonstrating that chloroacetanilides target fatty acid metabolism in target and nontarget organisms alike.

Though each herbicide targeted similar biochemical pathways in earthworms (i.e., fatty acid metabolism and β-oxidation, carbohydrate metabolism, TCA cycle intermediates, amino acid metabolism), the metabolites impacted in those pathways varied by compound. Significantly, we observed different metabolic profiles between earthworms exposed to racemic and S-metolachlor. Metolachlor chirality greatly affects herbicidal activity, and herein, we also detected differences in the metabolic profiles of exposure in a nontarget organism. These results suggest that stereospecific toxicant exposure can induce different metabolic profiles or phenotypes and that the effects of a structurally similar class of compounds may be discerned in the environment, supporting the utility of earthworm coelomic fluid in monitoring environmental insults.

4.5. References

1. Edwards, C. A.; Bohlen, P. J., *Biology and ecology of earthworms*. 3 ed.; Chapman & Hall: London, UK, 1996; p 11-12.

2. Simpson, M. J.; McKelvie, J. R., Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. *Anal. Bioanal. Chem.* **2009,** *394* (1), 137-149.

3. Lankadurai, B. P.; Nagato, E. G.; Simpson, M. J., Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environ. Rev.* **2013,** *21* (3), 180-205.

4. Fuerst, E. P., Understanding the mode of action of the chloroacetamide and thiocarbamate herbicides. *Weed Technol.* **1987,** *1* (4), 270-277.

5. Boger, P., Mode of action for chloroacetamides and functionally related compounds. *J. Pestic. Sci.* **2003,** *28* (3), 324-329.

6. US EPA Basic information about alachlor in drinking water. http://water.epa.gov/drink/contaminants/basicinformation/alachlor.cfm (accessed May 20, 2015).

7. Morton, M. D.; Walters, F. H.; Aga, D. S.; Thurman, E. M.; Larive, C. K., Nuclear magnetic resonance identification of new sulfonic acid metabolites of chloroacetanilide herbicides. *J. Agric. Food Chem.* **1997,** *45* (4), 1240-1243.

8. Graham, W. H.; Graham, D. W.; deNoyelles, F.; Smith, V. H.; Larive, C. K.; Thurman, E. M., Metolachlor and alachlor breakdown product formation patterns in aquatic field mesocosms. *Environ. Sci. Technol.* **1999,** *33* (24), 4471-4476.

9. Aga, D. S.; Thurman, E. M., Formation and transport of the sulfonic acid metabolites of alachlor and metolachlor in soil. *Environ. Sci. Technol.* **2001,** *35* (12), 2455-2460.

10. EU Pesticide Database S-metolachlor. http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.detail&language=EN&selectedID=1855 (accessed September 11, 2017).

11. USGS Estimated annual agricultural pesticide use. https://water.usgs.gov/nawqa/pnsp/usage/maps/compound_listing.php (accessed September 11, 2017).

12. Guo, J.; Li, Z.; Ranasinghe, P.; Bonina, S.; Hosseini, S.; Corcoran, M. B.; Smalley, C.; Kaliappan, R.; Wu, Y.; Chen, D.; Sandy, A. L.; Wang, Y.; Rockne, K. J.; Sturchio, N. C.; Giesy, J. P.; Li, A., Occurrence of atrazine and related compounds in sediments of upper great lakes. *Environ. Sci. Technol.* **2016,** *50* (14), 7335-7343.

13. Bradley, P. M.; Journey, C. A.; Romanok, K. M.; Barber, L. B.; Buxton, H. T.; Foreman, W. T.; Furlong, E. T.; Glassmeyer, S. T.; Hladik, M. L.; Iwanowicz, L. R.; Jones, D. K.; Kolpin, D. W.; Kuivila, K. M.; Loftin, K. A.; Mills, M. A.; Meyer, M. T.; Orlando, J. L.; Reilly, T. J.; Smalling, K. L.; Villeneuve, D. L., Expanded target-chemical analysis reveals extensive mixed-organic-contaminant exposure in u.S. Streams. *Environ. Sci. Technol.* **2017,** *51* (9), 4792-4802.

14. Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van der Vat, B. J. C.; Jellema, R. H., Fusion of mass spectrometry-based metabolomics data. *Anal. Chem.* **2005,** *77* (20), 6729-6736.

15. Boccard, J.; Rutledge, D. N., A consensus orthogonal partial least squares discriminant analysis (opls-da) strategy for multiblock omics data fusion. *Anal. Chim. Acta* **2013,** *769*, 30-39.

16. Florence, M.; Guillaume, M.; Philippe, M.; Estelle, D.; Lucie, B.; Horst, S.; Jean-Luc, W.; Serge, R.; Julien, B., Integrating metabolomic data from multiple analytical platforms for a comprehensive characterisation of lemon essential oils. *Flavour and Fragrance Journal* **2015,** *30* (2), 131-138.

17. Marshall, D. D.; Lei, S.; Worley, B.; Huang, Y.; Garcia-Garcia, A.; Franco, R.; Dodds, E. D.; Powers, R., Combining di-esi–ms and nmr datasets for metabolic profiling. *Metabolomics* **2015,** *11* (2), 391-402.

18. Griffith, C. M.; Williams, P. B.; Tinoco, L. W.; Dinges, M. M.; Wang, Y.; Larive, C. K., 1h nmr metabolic profiling of earthworm (eisenia fetida) coelomic fluid, coelomocytes, and tissue: Identification of a new metabolite – malylglutamate. *J. Proteome Res.* **2017,** *16* (9), 3407-3417.

19. OECD. Test no. 207: Earthworm, acute toxicity tests 1984. /content/book/9789264070042-en http://dx.doi.org/10.1787/9789264070042-en.

20. Wang, Y. H.; Wu, S. G.; Chen, L. P.; Wu, C. X.; Yu, R. X.; Wang, Q.; Zhao, X. P., Toxicity assessment of 45 pesticides to the epigeic earthworm eisenia fetida. *Chemosphere* **2012,** *88* (4), 484-491.

21. Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K., Comparison of gc-ms and nmr for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12* (2), 898-909.

22. Hummel, J.; Strehmel, N.; Bölling, C.; Schmidt, S.; Walther, D.; Kopka, J., Mass spectral search and analysis using the golm metabolome database. In *The handbook of plant metabolomics*, Wiley-VCH Verlag GmbH & Co. KGaA: 2013; pp 321-343.

23. Wheelock, A. M.; Wheelock, C. E., Trials and tribulations of 'omics data analysis: Assessing quality of simca-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* **2013,** *9* (11), 2589-2596.

24. Larson, M. G., Analysis of variance. *Circulation* **2008,** *117* (1), 115-121.

25. Ialongo, C., Understanding the effect size and its measures. *Biochem. Med.* **2016,** *26* (2), 150-163.

26. Bundy, J. G.; Osborn, D.; Weeks, J. M.; Lindon, J. C.; Nicholson, J. K., An nmr-based metabonomic approach to the investigation of coelomic fluid biochemistry in earthworms under toxic stress. *FEBS Lett.* **2001,** *500* (1-2), 31-35.

27. Yuk, J.; Simpson, M. J.; Simpson, A. J., Coelomic fluid: A complimentary biological medium to assess sub-lethal endosulfan exposure using h-1 nmr-based earthworm metabolomics. *Ecotoxicology* **2012,** *21* (5), 1301-1313.

28. Mudiam, M. K. R.; Ch, R.; Saxena, P. N., Gas chromatography-mass spectrometry based metabolomic approach for optimization and toxicity evaluation of earthworm sub-lethal responses to carbofuran. *PLoS One* **2013,** *8* (12), 13.

29. Ch, R.; Singh, A. K.; Pandey, P.; Saxena, P. N.; Reddy Mudiam, M. K., Identifying the metabolic perturbations in earthworm induced by cypermethrin using gas chromatography-mass spectrometry based metabolomics. *Sci. Rep.* **2015,** *5*, 15674.

30. Gillis, J. D.; Price, G. W.; Prasher, S., Lethal and sub-lethal effects of triclosan toxicity to the earthworm eisenia fetida assessed through gc-ms metabolomics. *J. Hazard. Mater.* **2017,** *323*, 203-211.

31. Jones, O. A. H.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L., A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm lumbricus rubellus. *Chemosphere* **2008,** *71* (3), 601-609.

32. Baylay, A. J.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L.; Swain, S.; Sturzenbaum, S.; Jones, O. A. H., A metabolomics based test of independent action and concentration addition using the earthworm lumbricus rubellus. *Ecotoxicology* **2012,** *21* (5), 1436-1447.

33. McKelvie, J. R.; Yuk, J.; Xu, Y. P.; Simpson, A. J.; Simpson, M. J., H-1 nmr and gc/ms metabolomics of earthworm responses to sub-lethal ddt and endosulfan exposure. *Metabolomics* **2009,** *5* (1), 84-94.

34. An, P. N. T.; Yamaguchi, M.; Fukusaki, E., Metabolic profiling of drosophila melanogaster metamorphosis: A new insight into the central metabolic pathways. *Metabolomics* **2017,** *13* (3), 29.

35. Rosenthal, J. A., Qualitative descriptors of strength of association and effect size. *J. Soc. Serv. Res.* **1996,** *21* (4), 37-59.

36. Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B., Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **2014,** *10* (3), 361-374.

37. Vaz, F. M.; Wanders, R. J. A., Carnitine biosynthesis in mammals. *Biochem. J.* **2002,** *361* (Pt 3), 417-429.

38. Wünschiers, R.; Jahn, M.; Jahn, D.; Schomburg, I.; Peifer, S.; Heinzle, E.; Burtscher, H.; Garbe, J.; Steen, A.; Schobert, M.; Oesterhelt, D.; Wachtveitl, J.; Chang, A., Metabolism. In *Biochemical pathways*, John Wiley & Sons, Inc.: 2012; pp 37-209.

39. Liebeke, M.; Bundy, J. G., Biochemical diversity of betaines in earthworms. *Biochem. Bioph. Res. Co,* **2013,** *430* (4), 1306-1311.

40. Rhee, H. J.; Kim, E.-J.; Lee, J. K., Physiological polyamines: Simple primordial stress molecules. *J. Cell Mol. Med.* **2007,** *11* (4), 685-703.

41. Gil, F. N.; Gonçalves, A. C.; Jacinto, M. J.; Becker, J. D.; Viegas, C. A., Transcriptional profiling in saccharomyces cerevisiae relevant for predicting alachlor mechanisms of toxicity. *Environ. Toxicol. Chem.* **2011,** *30* (11), 2506-2518.

42. Counihan, J. L.; Duckering, M.; Dalvie, E.; Ku, W.-m.; Bateman, L. A.; Fisher, K. J.; Nomura, D. K., Chemoproteomic profiling of acetanilide herbicides reveals their role in inhibiting fatty acid oxidation. *ACS Chem. Biol.* **2017,** *12* (3), 635-642.

43. Szewczyk, R.; Sobon, A.; Slaba, M.; Dlugonski, J., Mechanism study of alachlor biodegradation by paecilomyces marquandii with proteomic and metabolomic methods. *J. Hazard. Mater.* **2015,** *291*, 52-64.

44. Xu, H. D.; Wang, J. S.; Li, M. H.; Liu, Y.; Chen, T.; Jia, A. Q., H-1 nmr based metabolomics approach to study the toxic effects of herbicide butachlor on goldfish (carassius auratus). *Aquat. Toxicol.* **2015,** *159*, 69-80.

45. Xiao, N. W.; Song, Y.; Ge, F.; Liu, X. H.; Ou-Yang, Z. Y., Biomarkers responses of the earthworm eisenia fetida to acetochlor exposure in oecd soil. *Chemosphere* **2006,** *65* (6), 907-912.

46. Muthukaruppan, G.; Janardhanan, S.; Vijayalakshmi, G. S., Sublethal toxicity of the herbicide butachlor on the earthworm perionyx sansibaricus and its histological changes. *J. Soils Sediments* **2005,** *5* (2), 82-86.

47. Blaser, H. U.; Buser, H. P.; Coers, K.; Hanreich, R.; Jalett, H. P.; Jelsch, E.; Pugin, B.; Schneider, H. D.; Spindler, F.; Wegmann, A., The chiral switch of metolachlor: The development of a large-scale enantioselective catalytic process. *Chimia* **1999,** *53* (6), 275-280.

48. Xu, D. M.; Wen, Y. Z.; Wang, K. X., Effect of chiral differences of metolachlor and its (s)-isomer on their toxicity to earthworms. *Ecotox. Environ. Safe.* **2010,** *73* (8), 1925-1931.

49. Wu, J. R.; Hwang, I. T.; Hatzios, K. K., Effects of chloroacetanilide herbicides on membrane fatty acid desaturation and lipid composition in rice, maize, and sorghum. *Pest. Biochem. Physiol.* **2000,** *66* (3), 161-169.

CHAPTER FIVE

Metabolite Biomarkers of Chlorothalonil Exposure in Earthworms, Coelomic Fluid, and Coelomocytes using ^{1}H NMR and targeted LC-MS

Acknowledgements: I would like to thank Dr. Jay Kirkwood and the UCR Metabolomics Core Facility for awarding this work with a Seed Grant and for conducting the LC-MS analyses.

Abstract

Earthworm (*Eisenia fetida*) metabolomics is established as a useful indicator of toxicant exposure. Extracts of whole-earthworms are most commonly used to measure metabolic perturbations, in addition to coelomic fluid (CF) which has been used on a more limited basis. Coelomocytes (CC) are free moving liver-like and immune cells found within earthworm CF, and the potential of this compartment has not been evaluated for its utility in earthworm metabolomics. In this study, earthworms were exposed to 18.5 and 37.0 mg/kg chlorothalonil, a commonly used fungicide that targets glutathione, for 14-days and the metabolic impact on earthworm CF, and CC extracts assessed using ^{1}H NMR and targeted LC-MS. CF was identified as the most sensitive matrix for measuring chlorothalonil exposure, where an increase in glutamine levels was the only biomarker observed at both doses. At the high dose, multiblocked-orthogonal partial least squares-discriminant analysis (MB-OPLS-DA) supported increased *N*-acetylserine and ophthalmic acid levels as additional biomarkers of exposure in the CF. These perturbations may indicate increased oxidative stress, although no changes in glutathione were observed in

any matrix. ADP ribose was observed in the high dose group of earthworm and CC extracts, suggesting that chlorothalonil induced DNA damage. This work highlights the importance of evaluating the impacts of stressors in whole earthworm and CF to determine the most sensitive matrix for biomonitoring.

5.1. Introduction

Applied at over 4.5×10^6 kg per year, chlorothalonil (Figure 1) is one of the most heavily used fungicides in the United States.[1] Chlorothalonil is a nonsystemic fungicide that rapidly conjugates and depletes glutathione levels and inhibits NAD^+ thiol-dependent glycolytic and respiratory enzymes.[2-3] These biochemical pathway targets are universal, and although chlorothalonil is not designed to impact earthworms, it is reasonable to suspect that chlorothalonil exposure also targets glutathione, thiol-dependent enzymes, or related systems in nontarget organisms. Earthworms are an ideal model species to evaluate soil health and ecotoxicity because they are representative of the terrestrial environment, sensitive to contaminants, ecological engineers, and an important food source.[4] Environmental metabolomics and systems toxicology are growing tools to assess environmental and ecological toxicity, and earthworms are a useful tool for environmental monitoring.[5-9] Metabolite analysis of earthworms is commonly assessed in whole earthworm extracts;[5-6, 10] for instance, sub-lethal exposure of atrazine reduced ATP synthesis, as observed in changes in ATP, arginine, betaine, fumarate, glutamate, glutamine, malate, maltose, *scyllo*-inositol, and threonine/lactate levels.[11]

Additionally, coelomic fluid (CF) is a biofluid within the body cavity (i.e., coelom) of the earthworm and is crucial for movement, excretion, immunity, and metabolism, and is a complementary metabolic pool to measure metabolic perturbations.[4, 10, 12-14] Recently, metabolic profiling of chloroacetanilide herbicides in CF revealed perturbations in carnitine biosynthesis, energy demand, and lipid metabolism, affecting pathways similar to the herbicidal mode of action.[15] Coelomocytes (CC) are free-moving liver-like and immune cells within CF that have not been used in metabolomic studies, but their metabolome has been profiled using ^{1}H NMR as described in Chapter 2;[10] however, CC have been used in other toxicity assessment methods such as detecting DNA damage by H_2O_2 with the Comet Assay or identifying riboflavin content as a biomarker of lead and zinc exposure.[16-17]

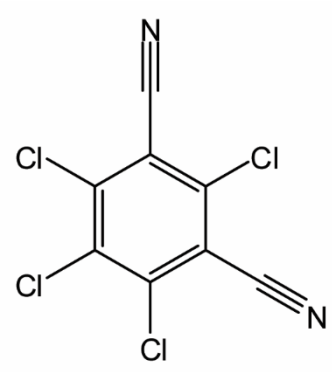
3. 

Figure 5.1. Chemical structure of chlorothalonil.

Metabolite measurements are primarily conducted with ^{1}H nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry. NMR is an inherently quantitative analytical tool with a large dynamic range capable of detecting metabolites in a less biased

fashion compared to other techniques.[6, 18-20] Metabolites are detected independently of pK_a, polarity, and solvent choice; however, NMR is limited to detecting micromolar or lower concentrations, using typical magnetic field strengths (500 – 700 MHz) and acquisition times (10 – 60 min). Mass spectrometry (MS) analysis offers increased sensitivity and an orthogonal method of metabolite detection. Liquid chromatography-mass spectrometry (LC-MS) is particularly useful for targeted analyses due to broad choices in solvents, column chemistry, and polarity that allow researchers to develop strategies to screen for particular metabolites or pathways; however, LC-MS is still well-suitable for untargeted approaches, although matrix effects and sample complexity can result in poor LC separations.[21] Nevertheless, LC-MS is superior in detecting more metabolites compared to other techniques due to its high sensitivity and chromatographic versatility (e.g., polarity, charge). One instrument cannot comprehensively measure every metabolite present in sample, therefore experiments that combine NMR and MS data can greatly expand metabolite coverage beyond what is possible using either technique alone.[15, 22-26]

To explore the metabolomic impact of chlorothalonil on earthworms, earthworms were exposed for 14-days in soil via the Organization for Economic Cooperation and Development (OECD) soil test guidelines at the application rate and half that value, 37 mg/kg and 18.5 mg/kg, respectively, which are equivalent to 13.8% and 6.9% of the reported LC_{50} of 268.5 mg/kg.[27-29] Coelomic fluid, coelomocytes, and whole earthworm samples were collected to compare metabolic impact within the three compartments. The outcome of these experiments is intended to identify the metabolite biomarkers and biochemical impacts of chlorothalonil exposure, and better understand which sample types,

whole earthworm, CF or CC, are more relevant for the measurement of metabolic perturbations of chlorothalonil using ^{1}H NMR and targeted LC-MS.

5.2. Experimental Procedures

5.2.1. Soil Exposure and Sample Preparation

To explore the metabolomic impact of chlorothalonil (Pestanal®, Sigma Aldrich, St. Louis, MO, USA) on earthworms, earthworms were exposed for 14-days in soil via the Organization for Economic Cooperation and Development (OECD) soil test guidelines at the application rate and half that value, 37.0 mg/kg and 18.5 mg/kg, respectively based on the soil dry weight.[27-28] Artificial soil was mixed in-house according to OECD instructions and contained 20% kaolin clay (Ward's Science, Rochester, NY, USA), 70% fine sand (Ward's Science, Rochester, NY, USA), and 10% Canadian sphagnum peat moss (Hoffman®, Lancaster, NY, USA). Soil was transferred at a rate of 50 g per worm to 1L beakers and dosed with chlorothalonil in acetone or acetone (control). The dosed soil was left in a fume hood to evaporate overnight, wetted to a 35% moisture content, and well-mixed prior to use.[30] Earthworms were removed from our laboratory culture, rinsed in ultrapure water, patted dry, and transferred to the beakers containing the soil. Sixty earthworms were used for this experiment, where twenty earthworms each where exposed to 0.0, 18.5, or 37.0 mg/kg chlorothalonil. Ten earthworms from each treatment were used for whole-earthworm analysis and the remaining ten were used for CF/CC analysis.

At the end of the 14-day exposure, earthworms were removed from the soil, rinsed with ultrapure water and patted dry. Earthworms intended for earthworm extracts were

flash-frozen and lyophilized, while earthworms intended for CF/CC were placed in a small petri dish containing 1 mL of 0.1% NaCl solution.[10, 12, 15] As described in Chapter 2 Section 2.2.2, CF/CC was extruded by apply a voltage across the worm (10x, < 1s each) using a 9V battery and a snap connector.[10, 12-13, 15] The CF/CC was transferred to a 2 mL Eppendorf tube, the petri dish was washed with 1 mL NaCl solution, and transferred to the same tube. The CF/CC was centrifuged at 400 *g*, 4°C for 10 min and the supernatant (i.e., CF) was transferred to a fresh tube. The CC pellet and CF were flash-frozen and lyophilized, and all samples were stored at -80°C until metabolite extraction.[31]

5.2.2. Metabolite Extraction

Prior to extraction, earthworm samples were homogenized by cryo-cooled bead-beating for three cycles of 10 s at 5 m/s (Omni International, Kennesaw, GA, US). Homogenized earthworms, CF, and CC samples were extracted using 2:2:1.8 Chloroform (Macron Fine Chemical, Center Valley, PA, USA):Fischer Scientific Optima Methanol (Fair Lawn, NJ, USA):Ultrapure Water.[32] All solvents used were ice-cold and samples were stored on ice when not in use. Aqueous solvents were first added to the tubes and vortexed at 2000 rpm, 4°C for 2 min on a ThermoMixer® C (Eppendorf, Hauppauge, NY, USA). The chloroform was added, and the samples were vortexed at 2000 rpm, 4°C for 2 min. The samples were centrifuged at 16000 *g*, 4°C for 20 min and left on ice for 10 min prior to dividing for NMR and LC-MS. A total of 1750 μL of solvent was used for each worm sample and 800 μL was transferred to a 1.5 mL Eppendorf tube, while a 1305 μL of solvent was used for CF and CC samples and 650 μL was transferred to a fresh tube. All

CF and CC samples were dried by speedvac overnight and stored in -80°C until NMR analysis. For LC-MS, 100 μL of extract was transferred to a conical insert (Fisher) and a QC sample was prepared for each matrix (i.e., earthworm, CF, CC) by transferring 3 μL of each sample into a single insert. LC-MS samples were immediately given to the UC Riverside Institute for Integrative Genome Biology (IIGB) Metabolomics Core for analysis.

5.2.3. ^{1}H NMR Metabolite Measurements and Preprocessing

The dried samples were reconstituted in 200 μL of 100 mM phosphate buffer (pD 7.4) containing 0.25 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid-d_6 (DSS-d_6) and 0.2 mM ethylenediaminetetraacetic acid-d_{16} (Cambridge Isotope Laboratories, Tewksbury, MA, USA), vortexed, and transferred to 3mm NE-H5/3-Br NMR tube (New Era Enterprises, Vineland, NJ, USA). ^{1}H NMR survey spectra were acquired using a Bruker Avance 600 MHz spectrometer operating at 599.88 MHz and equipped with a SmartProbe. Water suppression was conducted using the one-dimensional nuclear Overhauser effect spectroscopy (1D NOESY) with a 120 ms presaturation period (noesypr1d) during the 2 s relaxation delay. A spectral width of 11.8394 ppm was used, and 32768 complex data points were acquired using digital quadrature detection at 25°C. Spectra of earthworm extracts were acquired by co-adding 128 free-induction decays (FIDs), with16 dummy scans, and a 2.31 s acquisition time. Spectra of CF and CC samples were acquired by co-adding 512 FIDs, with 16 dummy scans, and a 2.31 s acquisition time.

Data was imported into MestReNova 12 (Santiago de Compostela, Spain) for preprocessing. Spectra were phased manually and referenced to DSS-d_{16} (0 ppm). FIDs were apodized by multiplication with an exponential function equivalent to 0.5 Hz and zero-filled to 131072. Peak fitting was conducted using a general Lorentzian peak shape, with a lower width constraint of 0.1 Hz, and upper constraint of 100 Hz, position constraint within ± 5%, maximum number of fine iterations of 100, and local minima filter of 5. The peak fitting results were exported into Excel (Microsoft, Redmond, WA, US) and assembled into a single spreadsheet. Resonance assignments were previously described,[10] and are summarized in Chapter 2.

5.2.4. LC-MS Metabolite Measurements and Preprocessing

A targeted LC-MS approach to detect and quantify polar, primary metabolites was employed using a Waters I-class UPLC system coupled to a TQ-XS triple quadrupole mass spectrometer. Separations were conducted on a ZIC-pHILIC column (2.1 x 150 mm, 5 µM) (EMD Millipore) using mobile phases (A) water with 15 mM ammonium bicarbonate adjusted to pH 9.6 with ammonium hydroxide and (B) acetonitrile. A 1 µL injection volume was used with a flow rate of 200 L/min and the column was held at 50° C. The gradient was as follows: 0 min, 90% B; 1.5 min, 90% B; 16 min, 20% B; 18 min, 20% B; 20 min, 90% B; 28 min.

The MS was operated in selected reaction monitoring mode. Source and desolvation temperatures were held at 150° C and 500° C, respectively. Nitrogen was used as the desolvation gas at 1000 L/h and cone gas at 150 L/h. Argon was used for the collision

gas and was set to 0.15 mL/min. The capillary voltage was 1 kV in positive ion mode and 2 kV in negative ion mode. System stability was monitored by periodically analyzing a quality control sample throughout the sample set. QC samples were prepared for each matrix by pooling together 3 μL of all sample extracts from a single matrix into a glass insert. To reduce systematic errors, samples were analyzed in random order. Peak integration was conducted with the open source software Skyline (University of Washington) by the staff of the UC Riverside Metabolomics Core Facility.[33]

5.2.5. Chemometric Analyses

Samples were normalized by earthworm dry-weight prior to chemometric analysis. For univariate analysis, samples were subjected to the Sharpiro-Wilks test of normality to assess skewness and Levene's test for homoscedasticity.[34] Metabolites were transformed using the logarithmic function to correct skewness. A one-way analysis of variance (ANOVA) was performed with Tukey's HSD *post hoc* test ($P \leq 0.05$) in IBM SPSS Statistics v24 (Armonk, NY, USA).[35] The effect size of statistically different metabolites was calculated using Glass' Δ.[36-37] Multivariate analysis was performed in SIMCA 14.1 (Sartorius Stedim Biotech, Malmö, Sweden). NMR and LC-MS were transformed using a logarithmic function, pareto scaled, and multiblocked using a block weight of 1/sqrt prior to analysis. Multiblocked-orthogonal partial least squares-discriminant analysis (MB-OPLS-DA) was used as the data reduction tool to visual grouping and assess metabolic perturbations. The number of predictive components and orthogonal components are reported [e.g, (1+1)]. R^2X describes the fit of the predictive component (x-axis) (e.g. R^2 =

1.0 means 100% of the data is explained), while R^2Y describes within group variation or the fit of the orthogonal component (y-axis).[38] Q^2 values ≥ 0.5 indicates good predictability of the model, describing the ability of the model to predict a new dataset.[38] Score plots were interpreted by plotting variance importance for the projection (VIP), which ranks the contribution of variables to the model, and p(corr) which is the loadings scaled around a correlation coefficient. Metabolites with VIP ≥ 1 and |p(corr)| ≥ 0.4 are considered most important to group separation. Receiver operating characteristics (ROC) curves were generated comparing control and high-dose following MB-OPLS-DA to further assess model predictability. An area under the curve (AUC) values of 1.0 means the model can correctly assign class 100% of the time.

5.3. Results and Discussion

^{1}H NMR and LC-MS were used to profile the impacts of chlorothalonil on the metabolome of *E. fetida*, and we detected 113 metabolites in earthworm extracts, 103 in CF extracts, and 84 in CC extracts (Figures 5.2 – 5.4). In earthworm extracts, 39 metabolites were detected using ^{1}H NMR, 97 were detected using LC-MS, and 23 were common in both methods (Figure 5.2). In CF extracts, 32 metabolites were detected using ^{1}H NMR, 82 using LC-MS, and 13 in common (Figure 5.3). Lastly, 29 metabolites were detected in the ^{1}H NMR spectra of CC extracts, 67 using LC-MS, and 12 in common (Figure 5.4). The ^{1}H NMR profiles of earthworms have been well characterized, detecting amino acids, betaine analogs, drilofensins, nucleotides, organic acid, polyamines, and sugars (Figure 5.2 – 5.4).[10, 39-47] LC-MS, to the best of our knowledge, has not been used

in earthworm metabolomic studies, although GC-MS has been widely utilized.[15, 22-24, 48-50] The targeted, polar LC-MS approach employed herein expanded metabolite coverage, providing insights into the effects on cofactors like S-adenosylmethionine (SAM) and flavin adenine dinucleotide (FAD); amino acid derivatives like argininosuccinic acid, glutathione, and 3,5-diiodatyrosine; and nucleotides and their derivatives such as CDP-choline, deoxycytidine, and xanthosine (Figures 5.2 – 5.4).

Once metabolites were identified, ANOVA using Tukey's HSD *post hoc* test ($P \leq 0.05$) and MB-OPLS-DA was conducted to assess metabolic perturbations. Tables 5.1 and 5.2 summarize the univariate results, showing the mean, standard deviation, *p*-values and Glass' Δ (i.e., effect size) of each compartment, where $|\Delta| \geq 0.8$ are considered strong and $|\Delta| \geq 1.3$ are considered very strong effect sizes.[36-37] Changes in metabolite levels are generally consistent between the NMR and LC-MS data; however, are a few discrepancies were observed. Several metabolites in the NMR spectra were not quantifiable due to resonance overlap or were present below limit of quantitation, such as aspartic acid in earthworm extracts and valine in CF extracts (Table 5.1). Overlapped resonances are also suspected to contribute to the lower significance for the glutamine increase in the low dose exposed group in the NMR data ($p = 0.043$) compared to LC-MS ($p = 0.008$). Several metabolites in the CF extracts showed near-significant increases in aminoadipate and nicotinamide adenine diphosphate (NAD^+) supported by pairwise comparisons of the low and high doses; however, valine showed near significance between the control and high-dose of LC-MS data (Table 5.1).

MB-OPLS-DA was selected as tool to visualize metabolic drivers between the different groups (Figure 5.5). Multiblocking (MB) allows for integration of the results obtained using orthogonal analytical instruments, like the NMR and MS datasets, into a single model to simplify data interpretation.[26, 51] Interpretation of the MB-OPLS-DA results was conducted by identifying variables with a |p(corr)| ≥ 0.4 (i.e., loadings scaled with a correlation coefficient) and variance important for the projection (VIP) ≥ 1 to identify the variables most important to group separation (Figure 5.5).[15, 38] Following MB-OPLS-DA, receiver operating characteristics (ROC) curves were generated to evaluate the diagnostic ability of the model (Figure 5.5).

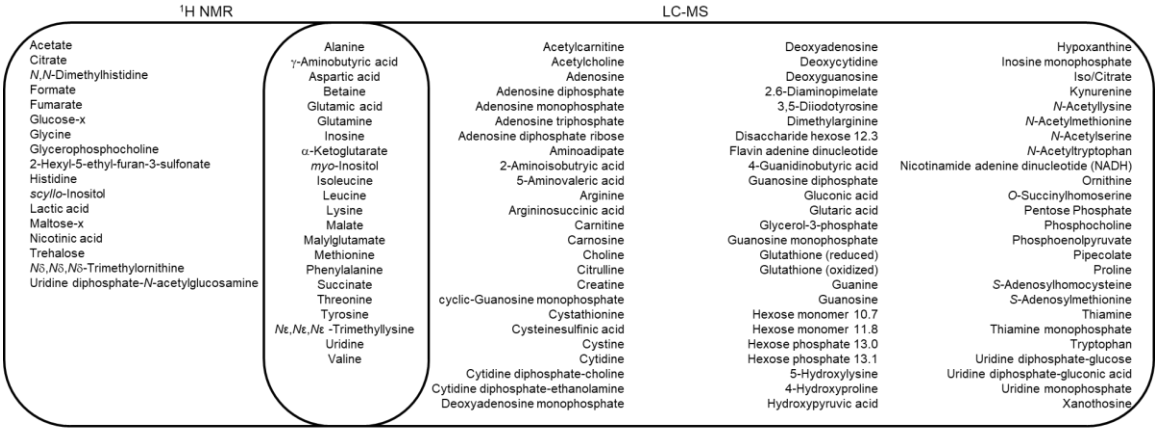


Figure 5.2. Venn diagram comparing metabolites detected with ^{1}H NMR (left), LC−MS (right), or both (center) in earthworm extracts.

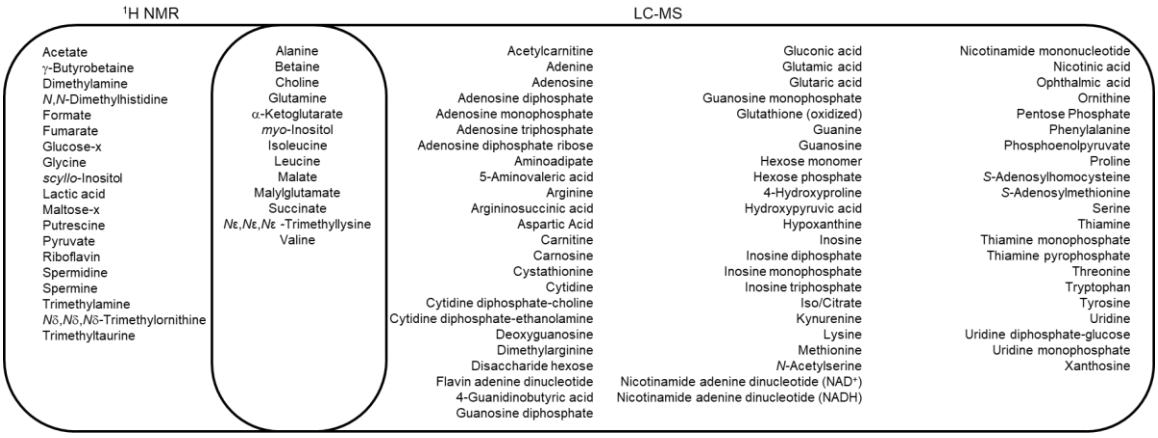


Figure 5.3. Venn diagram comparing metabolites detected with ^{1}H NMR (left), LC−MS (right), or both (center) in CF extracts.

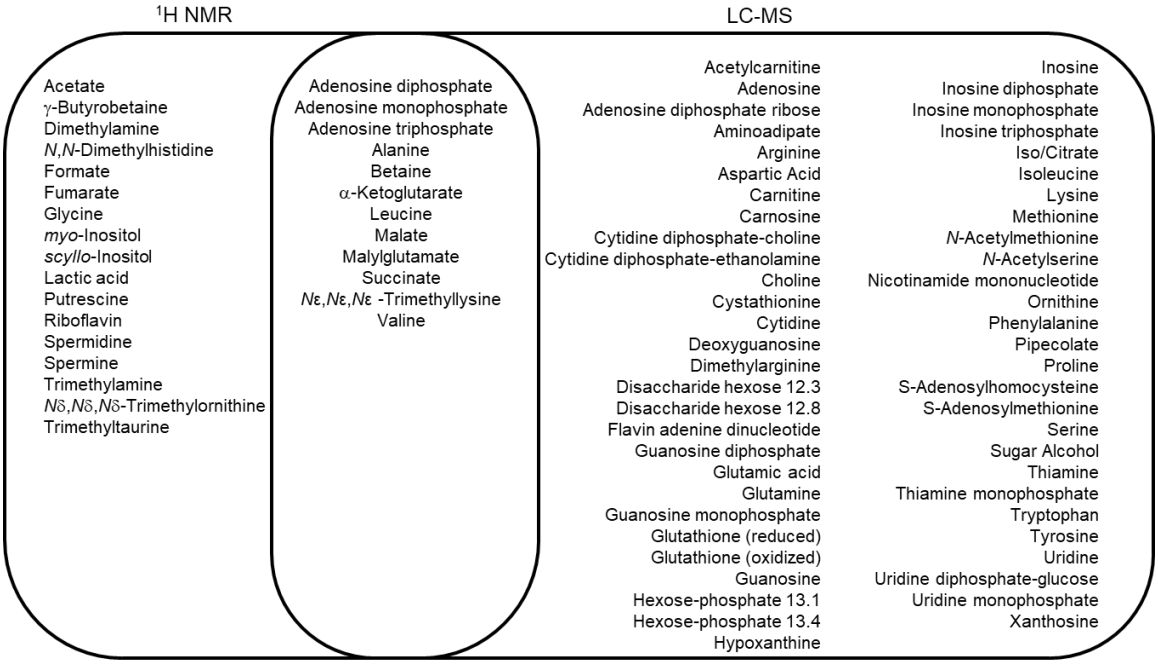


Figure 5.4. Venn diagram comparing metabolites detected with ^{1}H NMR (left), LC−MS (right), or both (center) in CC extracts.

Table 5.1. Statistically significant (P = 0.05) and Glass' Δ of metabolic changes in earthworms, CF, and CC extracts after 14-day exposure to 18.5 and 37.0 mg/kg chlorothalonil.

Chlorothalonil (mg/kg)	0.0-18.5				0.0-37.0				18.5-37.0			
	NMR		LC-MS		NMR		LC-MS		NMR		LC-MS	
	p	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ	*p*	Δ
Earthworm												
Acetylcholine	-	-			-	-	0.027	1.1	-	-	0.070	1.1
Adenosine Diphosphate (ADP) Ribose	-	-			-	-	0.024	1.0	-	-		
Adenosine Diphosphate (ADP)	-	-			-	-	0.022	1.0	-	-		
Adenosine Triphosphate (ATP)	-	-			-	-	0.028	1.4	-	-		
Aspartic Acid	NQ	NQ			NQ	NQ	0.024	1.4	NQ	NQ	0.057	0.8
Cytidine Diphosphate (CDP) - Ethanolamine	-	-			-	-	0.035	1.1	-	-		
Guanosine Diphosphate (GDP)	-	-			-	-	0.032	0.9	-	-		
Coelomic Fluid												
Adenosine Diphosphate (ADP) Ribose	-	-			-	-	0.082	1.2	-	-		
Aminoadipate	-	-			-	-	0.054	1.4	-	-	0.010	1.7
Glutamine	0.043	1.1	0.008	1.2	0.005	1.9	0.000	1.6		0.7		
N-Acetylserine	-	-			-	-	0.005	1.6	-	-	0.048	1.0
Nicotinamide Adenine Diphosphate (NAD^+)	-	-			-	-	0.053	1.2	-	-	0.028	1.0
Nitcotinamide Adenine Diphosphate (NADH)	-	-			-	-	0.025	1.2	-	-	0.024	1.0
Ophthalmic Acid	-	-			-	-	0.001	2.2	-	-	0.016	1.0
Tyrosine	-	-			-	-	0.037	1.1	-	-		
Valine	NQ	NQ			NQ	NQ	0.051	1.1	NQ	NQ		
Coelomocytes												
Adenosine Diphosphate (ADP) Ribose	-	-			-	-	0.017	1.2	-	-	0.013	1.9
Deoxyguanosine	-	-			-	-	0.039	1.4	-	-		

- = not detected

NQ = not quantifiable

Table 5.2. Mean and standard deviations of statistically significant (P = 0.05) metabolic changes in earthworms, CF, and CC extracts after 14-day exposure to 18.5 and 37.0 mg/kg chlorothalonil. Variables transformed with logarithmic function prior to ANOVA is noted.

Chlorothalonil (mg/kg)	0.0				18.5				37.0				
	NMR		LC-MS		NMR		LC-MS		NMR		LC-MS		
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	Transformation
Earthworm													
Acetylcholine	-	-	1.06E+06	4.85E+05	-	-	1.14E+06	3.91E+05	-	-	1.58E+06	3.83E+05	
ADP Ribose	-	-	5.49E+00	2.86E-01	-	-	5.58E+00	2.21E-01	-	-	5.78E+00	1.73E-01	log
ADP	-	-	5.61E+00	4.71E-01	-	-	5.93E+00	2.74E-01	-	-	6.10E+00	3.86E-01	log
ATP	-	-	1.79E+00	6.70E-01	-	-	2.18E+00	7.39E-01	-	-	2.74E+00	8.82E-01	log
Aspartic Acid	NQ	NQ	6.16E+00	1.31E-01	NQ	NQ	6.18E+00	1.91E-01	NQ	NQ	6.34E+00	1.02E-01	log
CDP-Ethanolamine	-	-	4.69E+00	1.64E-01	-	-	4.82E+00	1.83E-01	-	-	4.88E+00	1.03E-01	log
GDP	-	-	5.01E+00	2.09E-01	-	-	5.10E+00	1.45E-01	-	-	5.19E+00	7.17E-02	log
Coelomic Fluid													
ADP Ribose	-	-	3.79E+00	1.93E-01	-	-	3.96E+00	2.02E-01	-	-	4.02E+00	2.68E-01	log
Aminoadipate	-	-	4.95E+00	3.40E-01	-	-	4.80E+00	3.77E-01	-	-	5.45E+00	5.93E-01	log
Glutamine	7.96E-01	4.78E-01	6.87E+00	2.07E-01	1.32E+00	5.46E-01	7.11E+00	1.52E-01	1.72E+00	1.02E+00	7.20E+00	2.18E-01	log
N-Acetylserine	-	-	3.23E+00	2.37E-01	-	-	3.34E+00	2.65E-01	-	-	3.61E+00	2.77E-01	log
NAD^+	-	-	4.85E+00	1.83E-01	-	-	4.83E+00	2.46E-01	-	-	5.07E+00	1.41E-01	log
NADH	-	-	3.37E+00	2.01E-01	-	-	3.37E+00	2.48E-01	-	-	3.62E+00	1.23E-01	log
Ophthalmic Acid	-	-	3.39E+00	2.73E-01	-	-	3.54E+00	4.17E-01	-	-	3.98E+00	2.67E-01	log
Tyrosine	-	-	5.16E+00	2.40E-01	-	-	5.23E+00	2.53E-01	-	-	5.42E+00	1.58E-01	log
Valine	NQ	NQ	5.60E+00	2.01E-01	NQ	NQ	5.65E+00	1.97E-01	NQ	NQ	5.81E+00	1.78E-01	log
Coelomocytes													
ADP Ribose	-	-	3.16E+00	4.36E-01	-	-	3.15E+00	2.75E-01	-	-	3.68E+00	4.37E-01	log
Deoxyguanosine	-	-	2.93E+00	2.45E-01	-	-	3.06E+00	3.80E-01	-	-	3.28E+00	2.38E-01	log

- = not detected
NQ = not quantifiable

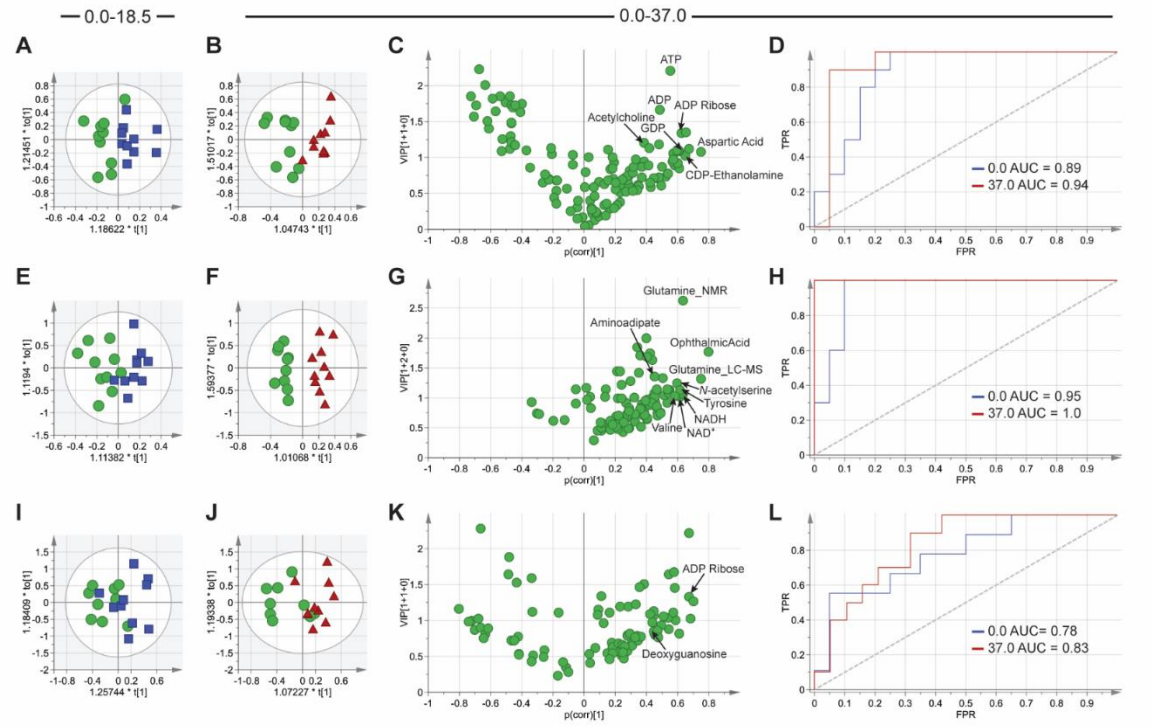


Figure 5.5. MB-OPLS-DA results of the metabolic impact of chlorothalonil exposure in earthworm (A-D), CF (E-H), and CC (J-L) extracts. Score plots (0.0 mg/kg = green circle, 18.5 mg/kg = blue square, and 37.0 mg/kg = red triangle), p(corr) vs. variance of importance (VIP) comparing 0.0-37.0 mg/kg annotated with metabolites that were statistically different in univariate analysis, and ROC curves of 0-37 mg/kg are shown. Results from earthworm extracts: (A) 0.0-18.5 mg/kg score plot [(1+1), R^2X=0.3, R^2Y=0.6, Q^2= -0.4]; (B) 0.0-37.0 mg/kg [(1+1), R^2X=0.4, R^2Y=0.8, Q^2=-0.1]; (C) p(corr) vs VIP plot; (D) ROC curve; Results from CF extract: (E) 0.0-18.5 mg/kg score plot [(1+1), R^2X=0.5, R^2Y=0.7, Q^2= -0.2]; (F) 0.0-37.0 mg/kg-dw score plot [(1+2+0), R2X=0.7, R2Y=0.2, Q2= 0.6]; (G) p(corr) vs VIP plot; (H) ROC curve; Results from CC extracts: (I) 0.0-18.5 mg/kg [(1+1), R^2X=0.6, R^2Y=0.4, Q^2= -0.4]; (J) 0.0-37.0 mg/kg score plot [(1+1), R^2X=0.6, R^2Y=0.5, Q^2= 0.1]; (K) p(corr) vs VIP plot; (L) ROC curve.

5.3.1. Metabolite Perturbations

Chlorothalonil perturbed metabolites involved in amino acid, energy, purine, and pyrimidine metabolism in earthworm, CF, and CC extracts (Table 5.1 – 5.2). Interestingly, ADP ribose was the only metabolite consistently affected in the high dose group of earthworms (p = 0.024) and CC (p = 0.017) extracts and increases in the CF extracts (p = 0.082) approached statistical significance (Table 5.1). CC were the least sensitive matrix to detect chlorothalonil exposure, with increases in only the levels of ADP ribose and deoxyguanosine observed. Several metabolite levels were altered in the high dose (37.0 mg/kg) exposed earthworm and CF extracts, but strikingly, glutamine was the only metabolite significantly affected in both the low and high dose groups, as observed in the CF extracts (Table 5.1 – 5.2) and was consistently detected by both ^{1}H NMR and LC-MS. Glutamine increases in CF extracts was the only significant perturbation detected by ^{1}H NMR in this matrix and therefore, LC-MS was more sensitive at detecting chlorothalonil metabolic perturbations in earthworms.

Several changes in metabolite levels were observed in the earthworm extracts of the high dose and all changes observed were considered strong or very strong effect sizes. Acetylcholine increased significantly with a strong effect size (Table 5.1 and Figure 5.6), which was supported by pairwise comparison of the low and high dose with p = 0.070. A similar trend was observed for aspartic acid, as shown in the box plots of Figure 5.6. Perturbations in purine and pyrimidine metabolites were observed with increases in ADP ribose, ADP, ATP, CDP-ethanolamine, and GDP levels. MB-OPLS-DA score plots showed reasonable grouping of the low and high dose samples compared with the control

(Figure 5.5AB); however, Q^2-values were below 0.5, indicating potential poor predictability and overfitting. Comparison of the p(corr) and VIP score of the control versus high dose (Figure 5.5C) indicated that all the metabolites statistically significant in the univariate analysis were associated with separation of the exposed group from the control with ADP, ATP, and aspartic acid among the metabolites most strongly driving the separation. Fumarate, inosine, and *N*,*N*-dimethylhistidine (all detected via NMR) were among the metabolites contributing most to the separation of the control and exposed groups. The predictability of the model was further evaluated using ROC curves (Figure 5.5D), where an area under the curve (AUC) of 0.89 was observed for the control group and 0.94 for the high dose group. In other words, the model can correctly predict group-classification of the control group 89% of the time (i.e., 11% false positives) and the high dose group 94% of the time (i.e., 6% false positive). The model is better at distinguishing members of the high dose group, which may indicate that biovariance in the data that is not group specific which leads to the insufficient Q^2-value.

CF was the most sensitive compartment to detect the impacts of chlorothalonil exposure on the metabolite profile. An increasing trend in glutamine levels (Figure 5.6) was observed at both doses, with glutamine being the only biomarker of earthworms exposed to 18.5 mg/kg chlorothalonil. Comparing the control and low dose groups, p = 0.043 and Δ = 1.1 were observed in the ^{1}H NMR data and p = 0.008 and Δ = 1.2 were calculated for the LC-MS data (Table 5.1). The higher p-value of the ^{1}H NMR data is suspected to be due to overlapped resonances that interfere with quantitation. Due to the similar trends observed in the box plots (Figure 5.6) and consistently strong effect sizes,

we believe the increase in glutamine levels is indeed real. The significance levels and effect sizes comparing the control and high dose groups were more similar, with $p = 0.005$, $\Delta = 1.9$ for ^{1}H NMR data and $p = 0.000$, $\Delta = 1.6$ for LC-MS data (Table 5.1). Other metabolite perturbations were observed in the high dose group, including *N*-acetylserine, aminoadipate, NAD^{+}, NADH, ophthalmic acid, tyrosine, and valine. Very significant increases and very strong effect sizes were observed for *N*-acetylserine ($p = 0.005$, $\Delta = 1.6$) and ophthalmic acid ($p = 0.001$, $\Delta = 2.2$) and further supported by the pairwise comparisons of low and high dose as seen in Figure 5.6 box plots. Increases in NAD^{+} levels were on the cusp of being statistically significant ($p = 0.053$), but with significance ($p = 0.028$) observed in low and high dose group comparisons. With similar trends observed for NADH, it seems likely that the change is real. The MB-OPLS-DA score plot comparing the control and low dose groups (Figure 5.5E) shows poor separation; however excellent separation was observed in the score plots comparing the control and high dose group (Figure 5.5F) and this is the only model to have a $Q^2 = 0.6$, meaning that it has good predictability. Figure 5.5G was used to ascertain the metabolite drivers of the separation, showing glutamine (NMR and LC-MS), ophthalmic acid, and *N*-acetylserine as the strongest drivers separating the high dose and control groups. Additionally, all other metabolites that were statistically significant in the univariate analysis were important for the separation of the high dose group from the control group (i.e., $p(corr) \geq 0.4$, $VIP \geq 1$). Other metabolites deemed important for the separation of the high dose group, include adenosine (LC-MS), betaine (NMR), dimethylarginine (LC-MS), *N*,*N*-dimethylhistidine (NMR), fumarate (NMR), glycine (NMR), hydroxypyruvic acid (LC-MS), kynurenine (LC-MS), leucine (LC-MS),

phenylalanine (LC-MS), threonine (LC-MS), and tryptophan (LC-MS). There were no metabolites that drove the separation of the control and exposed groups (i.e., p(corr) ≤ -0.4 and VIP ≥ 1), which may indicate larger biovariance in the control group. This could be further supported by the ROC curve (Figure 5.5H), where the AUC indicated that the model correctly classifies the control group 95% of the time and the high dose group 100% of the time, which is the strongest AUC on all models and matrices in this study.

CC extracts were the least sensitive matrix to detect the effects of chlorothalonil exposure. Increased ADP ribose (p = 0.017, Δ = 1.2) and deoxyguanosine (p = 0.034, Δ = 1.4) levels were observed in the high dose group (Table 5.1 and Figure 5.6). MB-OPLS-DA score plots revealed poor separation of the low and high dose groups from the control (Figure 5.5IJ) and the Q^2-value indicated low predictability. Due to the poor separation, Figure 5.5K cannot be used to conclusively suggest metabolites that drive separation, but ADP ribose is the only metabolite that is statistically significant and has a p(corr) ≥ 0.4 and VIP ≥ 1, indicating that it is a driver of the separation of the high dose and control groups. Poor predictability of the model is further supported by the ROC curves (Figure 5.5L) showing an AUC of 0.78 and 0.83 for control and high dose groups, respectively.

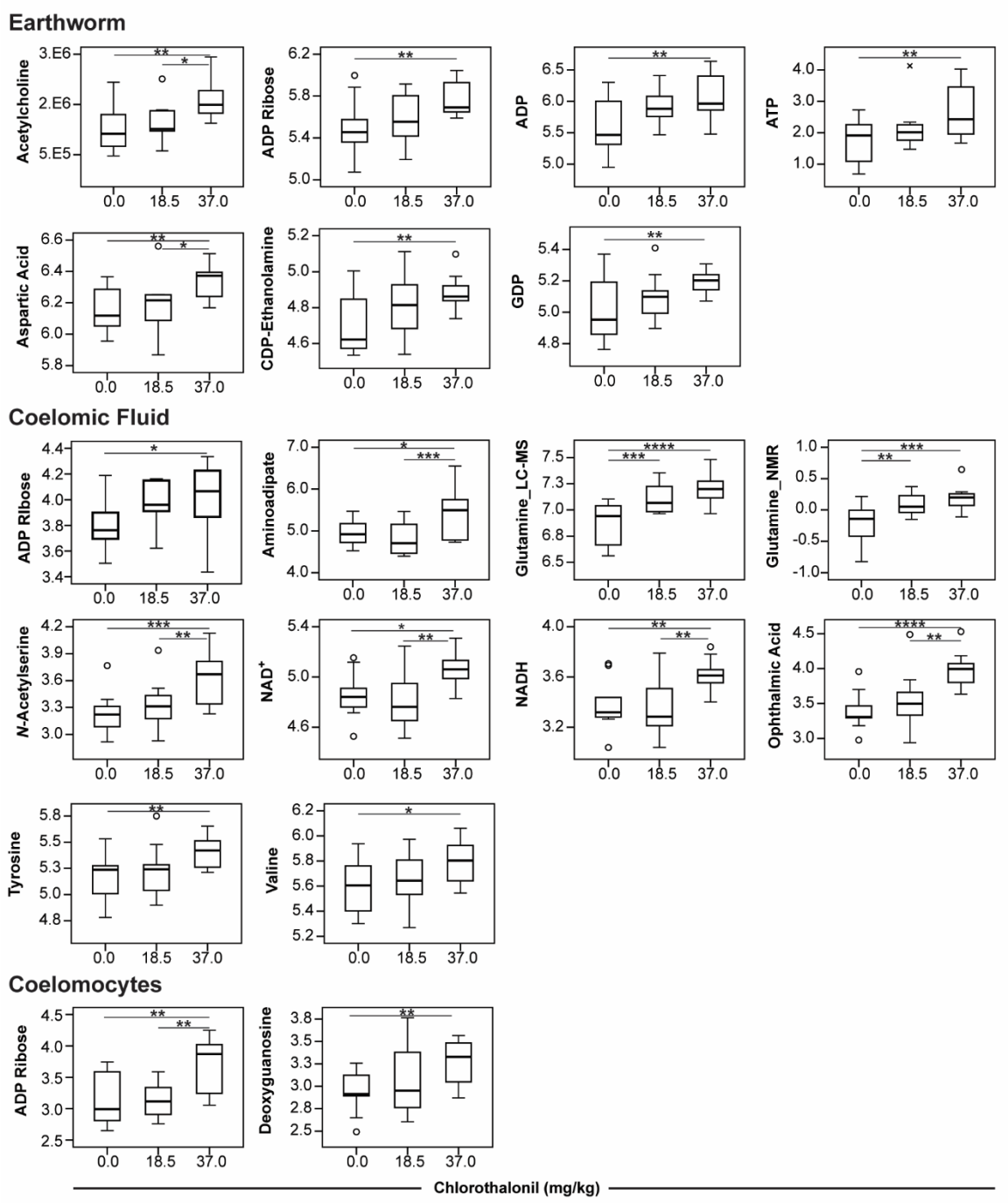


Figure 5.6. Box plots of statistically significant metabolite changes in earthworms, CF, and CC extracts exposed to chlorothalonil, as described in Table 5.1 and 5.2. ∗ = $P \leq 0.1$, ∗∗ = $P \leq 0.05$, ∗∗∗ = $P \leq 0.01$, ∗∗∗∗ = $P \leq 0.001$.

5.3.2. Matrix Comparisons and Biomarkers of Chlorothalonil Exposure

Major objectives of this study included determining (1) if the CC metabolome was useful in evaluating toxicant exposure, (2) which matrix was most sensitive to chlorothalonil exposure, and (3) what are the metabolite biomarkers of chlorothalonil exposure at environmentally relevant levels. Metabolites were affected in all matrices of earthworms exposed to soil dosed with 18.5 and 37.0 mg/kg chlorothalonil based on the soil dry weight. Compared with whole earthworm and CF extracts, the CC metabolites appeared to be less sensitive to chlorothalonil exposure due to greater biovariability. This is not surprising when one considers that CC are comprised of several cell types and it is likely that each sample contains a different relative proportion of cell populations. Therefore, it is difficult to conclude whether changes in metabolite levels are due exposure or differences in the composition of CC sample-to-sample. Flow cytometry methods have been established to separate subpopulations of earthworm CC prior to toxicity testing.[52] To more appropriately determine whether CC are useful in environmental metabolomics, CC would need to be separated following exposure and sample collection. This would increase experiment length and the separation process could affect cell metabolism,[53] and thus, it is probable that there is little advantage of using CC for biomonitoring in comparison to earthworm or CF extracts.

Perturbations of metabolite levels were detected in the earthworm and CF extracts following 14-day chlorothalonil exposure. Previously, earthworm and CF metabolite impacts were compared after exposure to endosulfan and endosulfan sulfate, which concluded that the two matrices are complementary for defining biomarkers and providing

insights into the mechanism of action.[13-14] Our study supports this notion, where we detected different metabolic perturbations in each matrix. Interesting, ADP ribose was the only consistent metabolite change observed in both matrices. CF was particularly useful for detecting chlorothalonil exposure, where glutamine was the only metabolite altered at 18.5 and 37.0 mg/kg chlorothalonil. Additionally, the greatest number of metabolic changes were measured in the CF. In conjunction with glutamine, *N*-acetylserine and ophthalmic acid increases could serve as useful metabolite biomarkers of earthworms exposed to 37.0 mg/kg chlorothalonil, as noted by univariate and multivariate analysis. Although fewer metabolites were perturbed in the earthworm extracts, several metabolite levels changed significantly in the high dose group. It is likely that the most sensitive matrix for monitoring toxicant exposures will depend on the compound of interest, and therefore, we propose profiling metabolites in both whole earthworm and CF to determine biomarkers of an exposure.

5.3.3. Potential Biochemical Targets

Observed metabolic perturbations in this study were different than the fungicidal mode of action of chlorothalonil. The main mechanism of action of chlorothalonil in fungi is to target glutathione and thiol-dependent enzymes. Though oxidized glutathione was detected in earthworm and CC extracts and reduced glutathione was detected in all matrices, statistically significant shifts in glutathione levels were not observed at the doses chosen and conditions used, granted other metabolite changes may be indicative of an increase in oxidative stress.[2] An increase in *N*-acetylserine levels was observed in the CF

extracts of earthworms exposed at the high dose. *N*-Acetylserine has been suggested as an inducer of cysteine biosynthesis and its increase could suggest an impact on cysteine or thiol metabolism.[54] Ophthalmic acid is an analog of glutathione that contains a 2-aminobutyrate moiety instead of cysteine, or in other words, a methyl group instead of thiol. We did not observe changes in oxidized glutathione (GSSG) and did not detected reduced glutathione (GSH) in CF extracts at the high dose, but levels of ophthalmic acid increased in the high dose in CF extracts (Figure 5.6). Intriguingly, ophthalmic acid increased with glutathione depletion in the livers of mice exposed to acetaminophen.[55] Soga, et al. [55] demonstrated that glutathione inhibits γ-glutamylcysteine synthase (GCS) under reducing conditions and as levels of glutathione are depleted, GCS is activated resulting in a increases in ophthalmic acid and the precursor γ-glutamyl-2-aminobutyrate. Under conditions and time point assessed in our experiments, chlorothalonil levels may have not been high enough to deplete glutathione levels but the observed increase in *N*-acetylserine and ophthalmic acid levels suggest that glutathione metabolism may have been affected.

Other metabolic perturbations demonstrated that chlorothalonil had an impact on amino acid, purine, and pyrimidine metabolism, and potentially induced DNA damage. ADP ribose is involved in the detection and repair of DNA damage, and was the only consistent metabolite change observed.[56] Poly ADP ribose polymerase (PARP) detects single strand breaks and initiates the production of poly-ADP ribose chains to signal DNA repair. Chlorothalonil-induced DNA damaged has been observed in *in vitro* studies with human peripheral blood lymphocytes and Chinese hamster ovary, and rat livers.[57-59] The

observed increase in ADP ribose in our study may suggest increased DNA damage and repair in earthworms exposed to chlorothalonil. PARP is NAD^+-dependent, which could be further supported by the increase in NAD^+ and NADH in the CF extracts. Chlorothalonil, in addition, targets NAD^+ dependent thiol glycolytic and respiratory enzymes which could also explain the increased NAD^+ and NADH levels.[2] Increases in ADP and ATP were also observed in the earthworm extracts, demonstrating that chlorothalonil exposure may have increased in energy consumption. Further research is warranted to determine the effects of chlorothalonil on glutathione metabolism and DNA damage in earthworms.

5.4. Conclusions

This study sought to evaluate metabolomic impacts of chlorothalonil, at environmentally relevant concentrations, in earthworm, coelomic fluid, and coelomocyte extracts and determine the utility of each matrix. Most metabolite changes were observed in earthworm and CF extracts, and we propose that in future studies both matrices be evaluated to determine biomarkers and stressor mode of action. Few metabolite changes in CC extracts were observed, which we postulate is due sample-to-sample coelomocyte subpopulation differences.

Metabolic profiling with orthogonal instrumentation broadens metabolic coverage, which was further supported by this research. Previously, we found that GC-MS increased coverage of CF metabolome, giving insights into fatty acids and other low-level metabolites (Chapter 4).[15] The targeted LC-MS method employed herein detected a larger

number metabolites compared with either ^{1}H NMR or GC-MS and was instrumental in detecting biomarkers of chlorothalonil exposure. These results suggest that LC-MS should be employed more often in earthworm metabolomic studies. Orthogonal instrumental methods also provide a means to confirm biomarkers and was useful in this study to confirm glutamine as a biomarker of chlorothalonil exposure.

Environmental metabolomic studies using environmentally relevant concentrations are essential to identifying relevant biomarkers of exposure. Identifying exposure markers can be challenging, since bioavailability of toxicants can change depending on environmental conditions. For instance, the bioavailability of chlorothalonil in soil can change depending on the organic matter content. Following OECD guidelines, a 10% organic matter content was chosen for this study, which is 5% higher than the soil used to determine the chlorothalonil LC_{50} for earthworms (268.5 mg/kg).[27, 29] Nevertheless, this study was able to determine that glutamine increases in earthworm CF following 14-day exposure to 37.0 mg/kg based on soil dry weight, an exposure equivalent to the application rate, and 18.5 mg/kg, the soil half-life concentration. Increasing *N*-acetylserine and ophthalmic acid levels, as seen for the 37 mg/kg dose, are additional metabolite signatures of chlorothalonil exposure. Although glutathione levels were not altered as hypothesized by the chlorothalonil mode of action in fungi, the changes in *N*-acetylserine and ophthalmic acid may indicate oxidative stress. Furthermore, increased levels of ADP ribose in earthworm and CC extracts might be representative of increased PARP activity to repair DNA damage induced by chlorothalonil. These results support the use of whole earthworm and CF extracts in monitoring chlorothalonil exposure at environmentally relevant levels.

5.5. References

1. USGS Estimated annual agricultural pesticide use. https://water.usgs.gov/nawqa/pnsp/usage/maps/compound_listing.php (accessed September 11, 2017).

2. Tillman, R. W.; Siegel, M. R.; Long, J. W., Mechanism of action and fate of the fungicide chlorothalonil (2,4,5,6-tetrachloroisophthalonitrile) in biological systems: I. Reactions with cells and subcellular components of saccharomyces pastorianus. *Pest. Biochem. Physiol.* **1973,** *3* (2), 160-167.

3. Van Scoy, A. R.; Tjeerdema, R. S., Environmental fate and toxicology of chlorothalonil. In *Reviews of environmental contamination and toxicology, vol 232*, Whitacre, D. M., Ed. Springer Int Publishing Ag: Cham, 2014; Vol. 232, pp 89-105.

4. Edwards, C. A.; Bohlen, P. J., *Biology and ecology of earthworms*. 3 ed.; Chapman & Hall: London, UK, 1996; p 11-12.

5. Simpson, M. J.; McKelvie, J. R., Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. *Anal. Bioanal. Chem.* **2009,** *394* (1), 137-149.

6. Lankadurai, B. P.; Nagato, E. G.; Simpson, M. J., Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environ. Rev.* **2013,** *21* (3), 180-205.

7. Sturla, S. J.; Boobis, A. R.; FitzGerald, R. E.; Hoeng, J.; Kavlock, R. J.; Schirmer, K.; Whelan, M.; Wilks, M. F.; Peitsch, M. C., Systems toxicology: From basic research to risk assessment. *Chem. Res. Toxicol.* **2014,** *27* (3), 314-329.

8. Gong, P.; Perkins, E. J., Earthworm toxicogenomics: A renewed genome-wide quest for novel biomarkers and mechanistic insights. *Appl. Soil Ecol.* **2016,** *104*, 12-24.

9. Zhang, X.; Xia, P.; Wang, P.; Yang, J.; Baird, D. J., Omics advances in ecotoxicology. *Environ Sci Technol* **2018,** *52* (7), 3842-3851.

10. Griffith, C. M.; Williams, P. B.; Tinoco, L. W.; Dinges, M. M.; Wang, Y.; Larive, C. K., 1h nmr metabolic profiling of earthworm (eisenia fetida) coelomic fluid, coelomocytes, and tissue: Identification of a new metabolite – malylglutamate. *J. Proteome Res.* **2017,** *16* (9), 3407-3417.

11. Dani, V. D.; Simpson, A. J.; Simpson, M. J., Analysis of earthworm sublethal toxic responses to atrazine exposure using (1) h nuclear magnetic resonance (nmr)-based metabolomics. *Environ Toxicol Chem* **2018,** *37* (2), 473-480.

12. Bundy, J. G.; Osborn, D.; Weeks, J. M.; Lindon, J. C.; Nicholson, J. K., An nmr-based metabonomic approach to the investigation of coelomic fluid biochemistry in earthworms under toxic stress. *FEBS Lett.* **2001,** *500* (1-2), 31-35.

13. Yuk, J.; Simpson, M. J.; Simpson, A. J., Coelomic fluid: A complimentary biological medium to assess sub-lethal endosulfan exposure using h-1 nmr-based earthworm metabolomics. *Ecotoxicology* **2012,** *21* (5), 1301-1313.

14. Yuk, J.; Simpson, M. J.; Simpson, A. J., 1-d and 2-d nmr-based metabolomics of earthworms exposed to endosulfan and endosulfan sulfate in soil. *Environ. Pollut.* **2013,** *175*, 35-44.

15. Griffith, C. M.; Morgan, M. A.; Dinges, M. M.; Mathon, C.; Larive, C. K., Metabolic profiling of chloroacetanilide herbicides in earthworm coelomic fluid using (1)h nmr and gc-ms. *J. Proteome Res.* **2018,** *17* (8), 2611-2622.

16. Plytycz, B.; Lis-Molenda, U.; Cygal, M.; Kielbasa, E.; Grebosz, A.; Duchnowski, M.; Andre, J.; Morgan, A. J., Riboflavin content of coelomocytes in earthworm (dendrodrilus rubidus) field populations as a molecular biomarker of soil metal pollution. *Environ. Pollut.* **2009,** *157* (11), 3042-3050.

17. Mincarelli, L.; Vischetti, C.; Craft, J.; Tiano, L., DNA damage in different eisenia andrei coelomocytes sub-populations after in vitro exposure to hydrogen peroxide. *Springerplus* **2016,** *5*.

18. Barding, G. A.; Salditos, R.; Larive, C. K., Quantitative nmr for bioanalysis and metabolomics. *Anal. Bioanal. Chem.* **2012,** *404* (4), 1165-1179.

19. Larive, C. K.; Barding, G. A.; Dinges, M. M., Nmr spectroscopy for metabolomics and metabolic profiling. *Anal. Chem.* **2015,** *87* (1), 133-146.

20. Giraudeau, P., Challenges and perspectives in quantitative nmr. *Magn. Reson. Chem.* **2017,** *55* (1), 61-69.

21. Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D., Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* **2014,** *87*, 12-25.

22. Jones, O. A. H.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L., A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm lumbricus rubellus. *Chemosphere* **2008,** *71* (3), 601-609.

23. McKelvie, J. R.; Yuk, J.; Xu, Y. P.; Simpson, A. J.; Simpson, M. J., H-1 nmr and gc/ms metabolomics of earthworm responses to sub-lethal ddt and endosulfan exposure. *Metabolomics* **2009,** *5* (1), 84-94.

24. Baylay, A. J.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L.; Swain, S.; Sturzenbaum, S.; Jones, O. A. H., A metabolomics based test of independent action and concentration addition using the earthworm lumbricus rubellus. *Ecotoxicology* **2012,** *21* (5), 1436-1447.

25. Barding, G. A.; Beni, S.; Fukao, T.; Bailey-Serres, J.; Larive, C. K., Comparison of gc-ms and nmr for metabolite profiling of rice subjected to submergence stress. *J. Proteome Res.* **2013,** *12* (2), 898-909.

26. Bhinderwala, F.; Wase, N.; DiRusso, C.; Powers, R., Combining mass spectrometry and nmr improves metabolite detection and annotation. *J. Proteome Res.* **2018,** *17* (11), 4017-4022.

27. OECD. Test no. 207: Earthworm, acute toxicity tests 1984. /content/book/9789264070042-en http://dx.doi.org/10.1787/9789264070042-en.

28. Chen, S.-K.; Edwards, C. A.; Subler, S., Effects of the fungicides benomyl, captan and chlorothalonil on soil microbial activity and nitrogen dynamics in laboratory incubations. *Soil Biol. Biochem.* **2001,** *33* (14), 1971-1980.

29. EC, Review report for the active substance chlorothalonil. European Commission Health and Consumer protection directorate General, D., D - Food Safety: Production and distribution chain, Unit D.3 - Chemicals, contaminants and pesticides, Ed. 2006; Vol. SANCO/4343/2000 final (revised).

30. Brown, S. A. E.; McKelvie, J. R.; Simpson, A. J.; Simpson, M. J., H-1 nmr metabolomics of earthworm exposure to sub-lethal concentrations of phenanthrene in soil. *Environ. Pollut.* **2010,** *158* (6), 2117-2123.

31. Lorenz, M. A.; Burant, C. F.; Kennedy, R. T., Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. *Anal. Chem.* **2011,** *83* (9), 3406-3414.

32. Bligh, E. G.; Dyer, W. J., A rapid method of total lipid extraction and purification. *Can. J. Biochem. Phys.* **1959,** *37* (8), 911-917.

33. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-968.

34. Levene, H., Robust tests for equality of variances. In *Contributions to probability and statistics: Essays in honor of harold hotelling*, Olkin, I., Ed. Stanford University Press: 1960; pp 278-292.

35. Tukey, J. W., Comparing individual means in the analysis of variance. *Biometrics* **1949,** *5* (2), 99-114.

36. Ialongo, C., Understanding the effect size and its measures. *Biochem. Med.* **2016,** *26* (2), 150-163.

37. Rosenthal, J. A., Qualitative descriptors of strength of association and effect size. *J. Soc. Serv. Res.* **1996,** *21* (4), 37-59.

38. Wheelock, A. M.; Wheelock, C. E., Trials and tribulations of 'omics data analysis: Assessing quality of simca-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* **2013,** *9* (11), 2589-2596.

39. Gibb, J. O. T.; Holmes, E.; Nicholson, J. K.; Weeks, J. M., Proton nmr spectroscopic studies on tissue extracts of invertebrate species with pollution indicator potential. *Comp. Biochem. Phys. B* **1997,** *118* (3), 587-598.

40. Gibb, J. O. T.; Svendsen, C.; Weeks, J. M.; Nicholson, J. K., 1h nmr spectroscopic investigations of tissue metabolite biomarker response to cu ii exposure in terrestrial invertebrates: Identification of free histidine as a novel biomarker of exposure to copper in earthworms. *Biomarkers* **1997,** *2* (5), 295-302.

41. Warne, M.; Lenz, E. M.; Osborn, D.; Weeks, J. M.; Nicholson, J. K., An nmr-based metabonomic investigation of the toxic effects of 3-trifluoromethyl-aniline on the earthworm eisenia veneta. *Biomarkers* **2000,** *5* (1), 56-72.

42. Bundy, J. G.; Spurgeon, D.; Svendsen, C.; Hankard, P.; Osborn, D.; Lindon, J.; Nicholson, J., Earthworm species of the genus eisenia can be phenotypically differentiated by metabolic profiling. *FEBS Lett.* **2002,** *521*, 115 - 120.

43. Lenz, E. M.; Weeks, J. M.; Lindon, J. C.; Osborn, D.; Nicholson, J. K., Qualitative high field h-1-nmr spectroscopy for the characterization of endogenous metabolites in earthworms with biochemical biomarker potential. *Metabolomics* **2005,** *1* (2), 123-136.

44. Brown, S. A. E.; Simpson, A. J.; Simpson, M. J., Evaluation of sample preparation methods for nuclear magnetic resonance metabolic profiling studies with eisenia fetida. *Environ. Toxicol. Chem.* **2008,** *27* (4), 828-836.

45. Liebeke, M.; Bundy, J. G., Biochemical diversity of betaines in earthworms. *Biochem. Bioph. Res. Co,* **2013,** *430* (4), 1306-1311.

46. Liebeke, M.; Strittmatter, N.; Fearn, S.; Morgan, A. J.; Kille, P.; Fuchser, J.; Wallis, D.; Palchykov, V.; Robertson, J.; Lahive, E.; Spurgeon, D. J.; McPhail, D.;

Takats, Z.; Bundy, J. G., Unique metabolites protect earthworms against plant polyphenols. *Nat Commun* **2015,** *6*, 7869.

47. Rochfort, S.; Wyatt, M. A.; Liebeke, M.; Southam, A. D.; Viant, M. R.; Bundy, J. G., Aromatic metabolites from the coelomic fluid of eisenia earthworm species. *Eur. J. Soil. Bio.* **2017,** *78*, 17-19.

48. Mudiam, M. K. R.; Ch, R.; Saxena, P. N., Gas chromatography-mass spectrometry based metabolomic approach for optimization and toxicity evaluation of earthworm sub-lethal responses to carbofuran. *PLoS One* **2013,** *8* (12), 13.

49. Ch, R.; Singh, A. K.; Pandey, P.; Saxena, P. N.; Reddy Mudiam, M. K., Identifying the metabolic perturbations in earthworm induced by cypermethrin using gas chromatography-mass spectrometry based metabolomics. *Sci. Rep.* **2015,** *5*, 15674.

50. Gillis, J. D.; Price, G. W.; Prasher, S., Lethal and sub-lethal effects of triclosan toxicity to the earthworm eisenia fetida assessed through gc-ms metabolomics. *J. Hazard. Mater.* **2017,** *323*, 203-211.

51. Marshall, D. D.; Lei, S.; Worley, B.; Huang, Y.; Garcia-Garcia, A.; Franco, R.; Dodds, E. D.; Powers, R., Combining di-esi–ms and nmr datasets for metabolic profiling. *Metabolomics* **2015,** *11* (2), 391-402.

52. Irizar, A.; Rivas, C.; García-Velasco, N.; Cerio, F. G.; Etxebarria, J.; Marigómez, I.; Soto, M., Establishment of toxicity thresholds in subpopulations of coelomocytes (amoebocytes vs. Eleocytes) of eisenia fetida exposed in vitro to a variety of metals: Implications for biomarker measurements. *Ecotoxicology* **2015,** *24* (5), 1004-1013.

53. Binek, A.; Rojo, D.; Godzien, J.; Rupérez, F. J.; Nuñez, V.; Jorge, I.; Ricote, M.; Vázquez, J.; Barbas, C., Flow cytometry has a significant impact on the cellular metabolome. *J. Proteome Res.* **2019,** *18* (1), 169-181.

54. Lynch, A. S.; Tyrrell, R.; Smerdon, S. J.; Briggs, G. S.; Wilkinson, A. J., Characterization of the cysb protein of klebsiella aerogenes: Direct evidence that n-acetylserine rather than o-acetylserine serves as the inducer of the cysteine regulon. *Biochem. J.* **1994,** *299* (1), 129-136.

55. Soga, T.; Baran, R.; Suematsu, M.; Ueno, Y.; Ikeda, S.; Sakurakawa, T.; Kakazu, Y.; Ishikawa, T.; Robert, M.; Nishioka, T.; Tomita, M., Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Bio. Chem.* **2006,** *281* (24), 16768-16776.

56. Krishnakumar, R.; Kraus, W. L., The parp side of the nucleus: Molecular actions, physiological outcomes, and clinical targets. *Molecular cell* **2010,** *39* (1), 8-24.

57. Lebailly, P.; Vigreux, C.; Godard, T.; Sichel, F.; Bar, E.; LeTalaër, J. Y.; Henry-Amar, M.; Gauduchon, P., Assessment of DNA damage induced in vitro by etoposide and two fungicides (carbendazim and chlorothalonil) in human lymphocytes with the comet assay. *Mutat. Res.* **1997,** *375* (2), 205-217.

58. Lodovici, M.; Casalini, C.; Briani, C.; Dolara, P., Oxidative liver DNA damage in rats treated with pesticide mixtures. *Toxicology* **1997,** *117* (1), 55-60.

59. Vigreux, C.; Poul, J. M.; Deslandes, E.; Lebailly, P.; Godard, T.; Sichel, F.; Henry-Amar, M.; Gauduchon, P., DNA damaging effects of pesticides measured by the single cell gel electrophoresis assay (comet assay) and the chromosomal aberration test, in chok1 cells. *Mutat. Res.* **1998,** *419* (1), 79-90.

CHAPTER SIX

Conclusions and Future Directions

6.1. Conclusions

This dissertation aimed to comprehensively assign metabolites in earthworm extracts using ^{1}H NMR, GC-MS, and LC-MS, and explore complementary matrices to whole-earthworm extracts, namely coelomic fluid (CF) and coelomocytes (CC), as bioindicators of environmental insult. Fifty-four metabolites were detected in earthworm extracts, 47 in CF, and 41 in CC using ^{1}H NMR, including amino acids, betaine analogs, organic acids, nucleotides, and polyamines (Chapters 2 and 3).[1] GC-MS was employed only with CF extracts where 44 metabolites were detected, increasing the coverage of fatty acids and low-abundance amino acids, organic acids, and derivatives (Chapter 4).[2] LC-MS expanded detection of primary metabolites, including amino acid derivatives, purines, pyrimidines, and signaling molecules (Chapter 5). The targeted LC-MS approach used detected 97 metabolites in earthworm extracts, 82 in CF, and 67 in CC. Current technologies cannot comprehensively detect all metabolites with a single instrument, and this work supports the use of orthogonal analytical instruments for metabolic profiling and confirmation of biomarkers.

Significantly, a newly identified metabolite, malylglutamate, found at approximately 7 μg/mg in earthworms was elucidated herein (Chapter 3).[1] In collaboration with Dr. David Martin's research group (UC Riverside), a synthesis scheme was

developed, and the absolute confirmation which was determined to be (−)-*β*-L-malyl-L-glutamate. Following identification, LC-MS was employed to explore the presence of malylglutamate in invertebrates closely related on the phylogenetic tree, where it was detected in redworms (*Eisenia fetida*), nightcrawlers (*Lumbricus sp.*), blackworms (*Lumbriculus variegatus*), black planaria (*Phagocata gracilis*), *Caenorhabditis elegans*, and water bears (*Hypsibius sp.*) at concentrations in the ng/mg - μg/mg range. It was not detected in white worms (*Enchytraeus sp*.), water fleas (*Daphnia Magna*), and mystery snails (*Pomacea bridgesi*), and potentially was detected in leeches (*Placobdella ornata*) and brown planaria (*Fugesia tigrine*).

Due to its high abundance in earthworms, malylglutamate is suspected to serve important biological functions and we postulated that malylglutamate could be a chelator, osmolyte, malate/glutamate store, and provided charge balance. Chelation by malylglutamate was observed using ^{1}H NMR where its resonances broadened and/or shifted downfield with increasing Mn^{2+}, Ca^{2+}, and Zn^{2+} concentrations, supporting the hypothesis that malylglutamate serves as chelator of micronutrients or a protectant from metal toxicity. Earthworms were exposed to cold stress, high water content, low water content, and high salt concentrations in soil to test the malylglutamate response as an osmolyte and store for malate/glutamate. Under these environmental stresses as well as exposure to the pesticides examined in this dissertation, no significant changes in malylglutamate levels were observed, so it is inconclusive whether malylglutamate has biological functions other than chelation.

In addition to metabolite assignment, this dissertation sought to explore the response of earthworm metabolism to environmental stressors. Chloroacetanilide herbicides (Chapter 4) were selected due to their extensive use and high mobility in the environment to further explore the response of CF to compounds in the same chemical class.[2] Earthworms were exposed to acetochlor, alachlor, butachlor, metolachlor, S-metolachlor, and propachlor for 2-days, following the OECD filter paper test guidelines.[3] CF was collected following exposure, and samples were analyzed using ^{1}H NMR and GC-MS. Metabolic perturbations in the earthworm CF were similar to the designed chloroacetanilide mechanism of action. Chloroacetanilide herbicides are a class of broad-spectrum, pre-emergent herbicides that target synthesis and elongation of very-long chain fatty acids.[4] The results suggested that chloroacetanilide herbicides disrupted fatty acids metabolism and increased energy demand, as seen by decreases in fatty acids levels and perturbations in metabolites involved in β-oxidation, carbohydrate, TCA cycle, and amino acid metabolism.

Chlorothalonil (Chapter 5) is one of the most heavily used fungicides worldwide. It targets glutathione and NAD^{+}-dependent thiol glycolytic and respiratory enzymes and was selected to compare how exposure of a toxicant affects metabolism in earthworm, CF, and CC extracts.[5] Environmentally relevant doses equivalent to the application and its half-life, 37.0 and 18.5 mg/kg soil dry weight, respectively, were chosen and worms were exposed in soil for 14-days.[3, 6] Earthworm, CF, and CC samples were collected and analyzed using ^{1}H NMR and LC-MS. Metabolic perturbations were detected in all matrices, with CC being the least sensitive matrix. CF extracts were the most sensitive

matrix to detect the effects of chlorothalonil exposure. Increased glutamine levels, detected by both instruments, was the only biomarker observed at both doses. Univariate and multivariate analysis also revealed *N*-acetylserine and ophthalmic acid as strong biomarker candidates of chlorothalonil exposure in the high dose group, which may indicate increased oxidative stress. ADP ribose was the only metabolite consistently affected increasing significantly in earthworm and CC extracts and nearly statistically significant in CF extracts. The increase in ADP ribose levels suggests an increase in poly ADP ribose polymerase in response to DNA damage induced by chlorothalonil. These results support metabolic profiling in whole earthworm and CF extracts to determine which matrix is most sensitive for detecting metabolite response to the stressor of interest.

6.2. Future Directions

6.2.1. Earthworm Metabolomics

Earthworm metabolomics has been studied as a possible indicator of environmental toxicity for over twenty years.[7-10] ^{1}H NMR has been the primarily analytical tool used for these studies, and still many unknowns remain the in the NMR spectra of all of the earthworm species commonly used for this research (e.g., *Eisenia fetida*, *Eisenia andrei*, *Eisenia veneta*, *Lumbricus terrestris*, *Aporrectodea caliginosa*). Metabolite annotation and profiling with GC-MS has been employed in several studies, and the targeted LC-MS analysis employed in Chapter 5 is the only known study thus far that has applied LC-MS in earthworm metabolic profiling. Several studies have used LC-MS to aid identification of suspected metabolites, for instance in Chapter 2 LC-MS was used to confirm the

assignment of *Nδ,Nδ,Nδ*-trimethylornithine, and Liebeke, et al. [11] used it to identify drilodefensins. Future research should explore untargeted LC-MS in earthworm metabolomics to expand metabolite coverage and aid identification of unknown metabolites. Furthermore, polar metabolites have primarily been targeted within the field, and only a couple recent papers have explored earthworm lipidomics.[12-13] Future research should consider incorporating lipid analysis into their workflows. This could help diagnose mechanism of action and identify biomarkers of environmental stressors. There is also a large knowledge gap of the lipids and other nonpolar metabolites that encompass the earthworm metabolome, and their identification could help aid understanding of earthworm biochemistry.

A large body of research has detected metabolite changes in response to toxicant exposure in eathworms.[9-10] Few investigations have stepped beyond detection of changes in metabolite levels and used biochemical approaches or other omics technologies to confirm predicted biochemical pathway perturbations and biomarkers of exposure. There is a need to take a more biochemical and broader omics approach within the field to not only confirm hypotheses but understand the utility of earthworm metabolomics in biomonitoring, including determining its sensitivity in comparison to other methods.

6.2.2. Malylglutamate

The identification of malylglutamate and exploration of its biological function is a significant contribution of this dissertation research to the field of earthworm metabolomics. In Chapter 3, malylglutamate was detected in several species close on the

phylogenetic tree. We detected malylglutamate in invertebrate species closely related to *E. fetida*, and the presence of malylglutamate in other organism, such as vertebrates, fungi, bacteria, plants, should be explored.

Malylglutamate was determined to be a chelator (Chapter 3) and further studies should be conducted to study malylglutamate concentration under heavy metal stress to see how its levels are affected. Furthermore, it is important to measure the binding affinities of malylglutamate with a variety of metal ions to determine if it is a general chelator or designed to specifically chelate micronutrient metals or defend against metal toxicity. Further studies are needed to explore what other biological functions malylglutamate serves in earthworms. Our results suggested that it may serve as a malate and glutamate and potentially provides charge balance; however, this needs to be more thoroughly investigated, perhaps by knocking out the genes that produce malylglutamate and then evaluate their fitness under environmental stress (e.g. salinity, temperature, water content, metals) to elucidate function.

Lastly, we hypothesized that malylglutamate is synthesized by the same family of enzymes as *N*-acetylaspartylglutamate (NAAG) and *β*-citrylglutamate.[14-15] Further research should be conducted to explore the presence and analogs of the RIMKLA and RIMKLB genes in earthworms, or *C. elegans* since their genome is annotated, to test this hypothesis. The synthetic pathway of malylglutamate can also be explored through the incorporation of stable isotopes to follow its flux, which could give insights into its metabolism and biochemical function.

6.3. References

1. Griffith, C. M.; Williams, P. B.; Tinoco, L. W.; Dinges, M. M.; Wang, Y.; Larive, C. K., 1h nmr metabolic profiling of earthworm (eisenia fetida) coelomic fluid, coelomocytes, and tissue: Identification of a new metabolite – malylglutamate. *J. Proteome Res.* **2017,** *16* (9), 3407-3417.

2. Griffith, C. M.; Morgan, M. A.; Dinges, M. M.; Mathon, C.; Larive, C. K., Metabolic profiling of chloroacetanilide herbicides in earthworm coelomic fluid using (1)h nmr and gc-ms. *J. Proteome Res.* **2018,** *17* (8), 2611-2622.

3. OECD. Test no. 207: Earthworm, acute toxicity tests 1984. /content/book/9789264070042-en http://dx.doi.org/10.1787/9789264070042-en.

4. Boger, P., Mode of action for chloroacetamides and functionally related compounds. *J. Pestic. Sci.* **2003,** *28* (3), 324-329.

5. Tillman, R. W.; Siegel, M. R.; Long, J. W., Mechanism of action and fate of the fungicide chlorothalonil (2,4,5,6-tetrachloroisophthalonitrile) in biological systems: I. Reactions with cells and subcellular components of saccharomyces pastorianus. *Pest. Biochem. Physiol.* **1973,** *3* (2), 160-167.

6. Chen, S.-K.; Edwards, C. A.; Subler, S., Effects of the fungicides benomyl, captan and chlorothalonil on soil microbial activity and nitrogen dynamics in laboratory incubations. *Soil Biol. Biochem.* **2001,** *33* (14), 1971-1980.

7. Gibb, J. O. T.; Holmes, E.; Nicholson, J. K.; Weeks, J. M., Proton nmr spectroscopic studies on tissue extracts of invertebrate species with pollution indicator potential. *Comp. Biochem. Phys. B* **1997,** *118* (3), 587-598.

8. Gibb, J. O. T.; Svendsen, C.; Weeks, J. M.; Nicholson, J. K., 1h nmr spectroscopic investigations of tissue metabolite biomarker response to cu ii exposure in terrestrial invertebrates: Identification of free histidine as a novel biomarker of exposure to copper in earthworms. *Biomarkers* **1997,** *2* (5), 295-302.

9. Lankadurai, B. P.; Nagato, E. G.; Simpson, M. J., Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environ. Rev.* **2013,** *21* (3), 180-205.

10. Simpson, M. J.; McKelvie, J. R., Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil. *Anal. Bioanal. Chem.* **2009,** *394* (1), 137-149.

11. Liebeke, M.; Strittmatter, N.; Fearn, S.; Morgan, A. J.; Kille, P.; Fuchser, J.; Wallis, D.; Palchykov, V.; Robertson, J.; Lahive, E.; Spurgeon, D. J.; McPhail, D.; Takats, Z.; Bundy, J. G., Unique metabolites protect earthworms against plant polyphenols. *Nat Commun* **2015,** *6*, 7869.

12. He, Z.; Wang, Y.; Zhang, Y.; Cheng, H.; Liu, X., Stereoselective bioaccumulation of chiral pcb 91 in earthworm and its metabolomic and lipidomic responses. *Environ. Pollut.* **2018,** *238*, 421-430.

13. Koelmel, J. P.; Jones, C. M.; Ulmer, C. Z.; Garrett, T. J.; Yost, R. A.; Schock, T. B.; Bowden, J. A., Examining heat treatment for stabilization of the lipidome. *Bioanalysis* **2018,** *10* (5), 291-305.

14. Collard, F.; Stroobant, V.; Lamosa, P.; Kapanda, C. N.; Lambert, D. M.; Muccioli, G. G.; Poupaert, J. H.; Opperdoes, F.; Van Schaftingen, E., Molecular identification of n-acetylaspartylglutamate synthase and β-citrylglutamate synthase. *J. Bio. Chem.* **2010,** *285* (39), 29826-29833.

15. Collard, F.; Vertommen, D.; Constantinescu, S.; Buts, L.; Van Schaftingen, E., Molecular identification of beta-citrylglutamate hydrolase as glutamate carboxypeptidase 3. *J. Bio. Chem.* **2011,** *286* (44), 38220-38230.

APPENDIX A

Acknowledgements: I would like to thank Dr. David Martin and Abigale Fecue for conducting the malylglutamate synthesis and optical rotation measurements for this work.

A.1. Malylglutamate Synthesis Procedures

^{1}H and ^{13}C NMR spectra were recorded on a Varian Inova 400 MHz spectrometer unless otherwise indicated and were internally referenced to residual solvent signal (note: D_2O was referenced at 4.79 ppm for ^{1}H and $CDCl_3$ referenced at 7.26 ppm for ^{1}H NMR and 77.16 ppm for ^{13}C NMR, respectively). Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), and coupling constant (Hz). Data for ^{13}C NMR are reported in terms of chemical shift and no special nomenclature is used for equivalent carbons. IR spectra were recorded on a Bruker Alpha FT-IR Spectrometer. High-resolution mass spectrometry data were recorded on an Agilent LCTOF instrument using direct injection of samples in dichloromethane into the electrospray source (ESI) with positive ionization. Optical rotation was measured on a Rudolph Research Analytical Autopol IV Automatic Polarimeter.

All reactions were carried out under an inert atmosphere of nitrogen in oven dried or flame dried glassware with magnetic stirring, unless otherwise noted. Solvents were dried by passage through columns of activated alumina. All starting materials were prepared according to known literature procedures or used as obtained from commercial

sources, unless otherwise indicated. Reactions were monitored by thin-layer chromatography (TLC) and carried out on 0.25 mm coated commercial silica gel plates (Analtech TLC Uniplates, F_{254} precoated glass plates with organic fluorescent binder) using UV light as visualizing agent and $KMnO_4$ and heat as a developing agent. Flash chromatography was performed on silica gel (Silicycle, SiliaFlash P60, 230-400 mesh).

A.1.1. Synthesis and Characterization of Peptide Coupling Partners

L-6, D-6, L-8, L-9, L-11, and D-11 (Figures 1 and 2) were synthesized according to known literature procedures.[1-4]

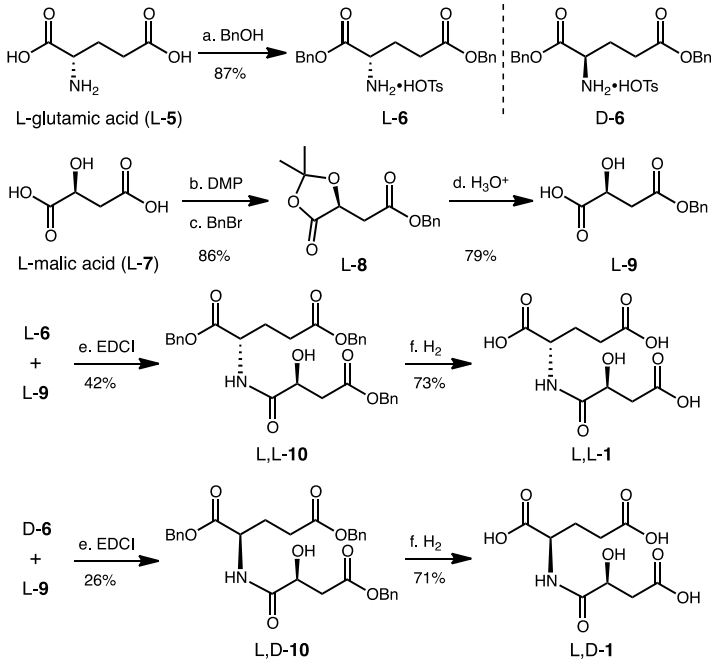


Figure 1. Synthesis of both diastereomers of the originally proposed structure **1**. a. BnOH, 1.22 equiv *p*-TsOH, toluene, rt. b. 2,2-dimethoxypropane (DMP), 1 mol % *p*-TsOH, CH_2Cl_2. c. BnBr, Cs_2CO_3, DMF. d. AcOH, H_2O, 60 ºC. e. EDCI, HOBt, NEt_3, DMF. f. H_2, 5 mol % Pd/C, EtOH.

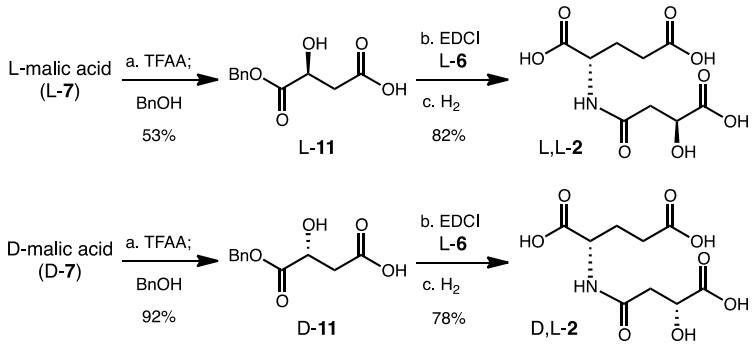
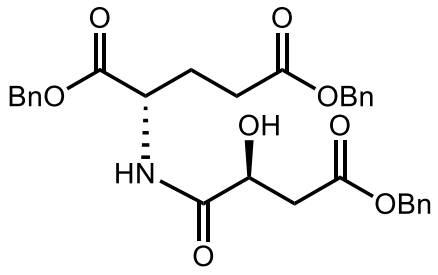


Figure 2. Synthesis of both diastereomers of the revised structure 2. a. TFAA, BnOH. b. EDCI, HOBt, NEt_3, DMF. c. H_2, 5 mol % Pd/C, EtOH.

A.1.2. Synthesis and Characterization of Benzylated Compounds

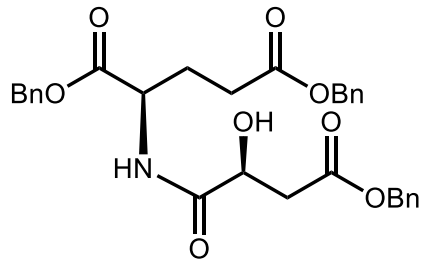
General Procedure for Peptide Coupling Reactions (A): To a flame dried flask equipped with a magnetic stir bar were sequentially added acid (1.0 eq), amine (1.2 eq), HOBT (1.2 eq), EDCI (2.0 eq) and DMF (60 mM). The vial was sealed with a septum and degassed. The reaction mixture was stirred for 10 minutes followed by addition of Et_3N (4.0 eq) dropwise. The reaction was stirred for 24 hours followed by extraction with EtOAc and dried over Na_2SO_4. The solvent was removed in vacuo and the crude residue was purified using silica gel chromatography to provide benzylated products as a white solid.



L,L-**10**

(*S*)-Dibenzyl 2-((*S*)-4-(benzyloxy)-2-hydroxy-4-oxobutanamido)pentanedioate (10)

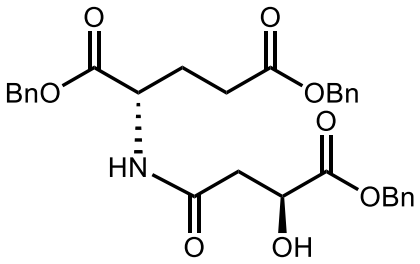
Prepared according to General Procedure A using acid L-9 (22 mg, 0.10 mmol) and amine L-6 (60 mg, 0.12 mmol) as the substrates. Reaction time: 24 h. The crude residue was purified by column chromatography on silica gel using EtOAc to afford L,L-10 as a white solid (22 mg, 42% yield). IR (film) 3388, 3063, 3032, 1735, 1731 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.98-2.10 (m, 1H), 2.20-2.34 (m, 1H), 2.35-2.50 (m, 2H), 2.82 (dd, *J* = 7.88, 17.44 Hz, 1H), 2.94 (dd, *J* = 3.88, 17.48 Hz, 1H), 3.60-3.66 (m, 1H), 4.40-4.42 (m, 1H), 4.64-4.72 (M, 1H), 5.10 (s, 2H), 5.13 (s, 2H), 5.16 (s, 2H), 7.27-7.39 (m, 15H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 172.6, 172.4, 172.2, 171.4, 135.9, 135.4, 135.3, 133.9, 128.8, 128.7, 128.7, 128.6, 128.5, 128.4, 128.4, 77.5, 77.2, 76.9, 72.2, 68.6, 67.5, 67.5, 67.0, 66.7, 51.6, 38.8, 38.4, 30.3, 29.3, 27.4, 27.3; HRMS (ESI) *m/z* calcd. for $C_{30}H_{31}NO_8$ $(M+H)^+$ 533.2044, found 533.2040.



L,D-**10**

(*R*)-Dibenzyl 2-((*S*)-4-(benzyloxy)-2-hydroxy-4-oxobutanamido)pentanedioate (10)

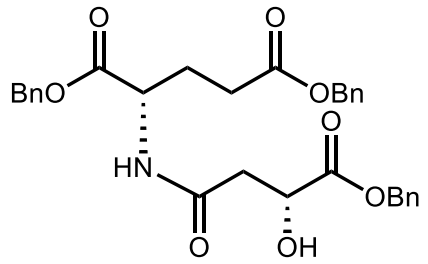
Prepared according to General Procedure A using acid L-9 (100 mg, 0.446 mmol) and amine D-6 (267 mg, 0.535 mmol). Reaction time: 24 h. The crude residue was purified by column chromatography on silica gel with EtOAc to afford L,D-10 as a white solid (63 mg, 26% yield). IR (film) 3389, 3033, 2948, 1731, 1715 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.00-2.10 (m, 1H), 2.20-2.33 (m, 1H), 2.35-2.50 (m, 2H), 2.74 (dd, *J* = 8.72, 17.44 Hz, 1H), 2.94 (dd, *J* = 3.6, 17.56 Hz, 1H), 3.65 (d, *J* = 4.92 Hz, 1H), 4.40-4.48 (m, 1H), 4.62-4.72 (m, 1H), 5.09 (s, 2H), 5.15 (s, 2H), 5.16 (s, 2H), 7.28-7.40 (m, 15H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 172.6, 172.0, 171.4, 164.9, 138.6, 135.9, 135.4, 133.9, 128.8, 128.7, 128.7, 128.6, 128.5, 128.5, 128.4, 128.0, 127.8, 77.5, 77.2, 76.9, 72.3, 68.6, 67.6, 67.3, 67.1, 66.7, 51.6, 38.3, 29.9, 27.4; HRMS (ESI) *m/z* calcd. for $C_{30}H_{31}NO_8$ $(M+H)^+$ 533.2044, found 533.2037.



L,L-S1

(*S*)-Dibenzyl 2-((*S*)-4-(benzyloxy)-3-hydroxy-4-oxobutanamido)pentanedioate

Prepared according to General Procedure A using acid L-9 (500 mg, 2.23 mmol) and amine L-6 (1.34 g, 2.67 mmol). Reaction time: 24 h. The crude residue was purified by column chromatography on silica gel using EtOAc to afford L,L-S1 as a white solid (868 mg, 73% yield). IR (film) 3360, 2915, 1745, 1715, 1170 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.94-2.06 (m, 1H), 2.16-2.27 (m, 1H), 2.33-2.48 (m, 1H), 2.62 (dd, *J* = 7.16, 15.32 Hz, 1H), 2.72 (dd, *J* = 3.68, 15.32 Hz, 1H), 3.61 (bs, 1H), 4.50 (dd, *J* = 3.72, 7.12 Hz, 1H), 4.63-4.70 (m, 1H), 5.09 (s, 2H), 5.15 (s, 2H), 5.20 (s, 2H), 6.56 (d, *J* = 7.84, 1H), 7.28-7.39 (m, 15H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 173.3, 172.8, 171.6, 170.0, 135.8, 135.2, 128.8, 128.7, 128.7, 128.6, 128.5, 128.5, 77.5, 77.2, 67.9, 67.8, 67.8, 67.6, 66.7, 51.9, 40.0, 31.0, 30.3, 27.3; HRMS (ESI) *m/z* calcd. for $C_{30}H_{31}NO_8$ $(M+H)^+$ 533.2044, found 533.2049.



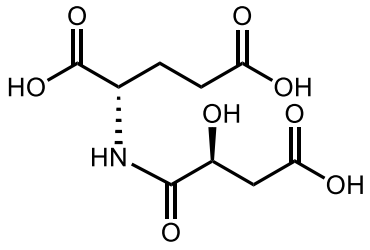
D,L-S2

(*S*)-Dibenzyl 2-((*R*)-4-(benzyloxy)-3-hydroxy-4-oxobutanamido)pentanedioate

Prepared according to General Procedure A using acid D-11 (400 mg, 1.78 mmol) and amine L-6 (1.07 g, 2.14 mmol). Reaction time: 24 h. The crude residue was purified by column chromatography on silica gel EtOAc to afford D,L-S2 as a white solid (683 mg, 72% yield). IR (film) 3352, 3065, 2853, 1733 cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.94-2.06 (m, 1H), 2.16-2.28 (m, 1H), 2.32-2.48 (m, 2H), 2.63 (dd, *J* = 7.12, 15.28 Hz, 1H), 2.72 (dd, *J* = 3.72, 15.32 Hz, 1H), 3.55 (d, *J* = 5.52 Hz, 1H), 4.47-4.53 (m, 1H), 4.62-4.70 (m, 1H), 5.09 (s, 2H), 5.15 (s, 2H), 5.21 (s, 2H), 6.53 (d, *J* = 7.92 Hz, 1H), 7.30-7.39 (m, 15H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 177.7, 173.3, 172.3, 170.4, 135.5, 135.1, 128.70, 128.6, 128.5, 128.4, 128.4, 128.3, 77.5, 77.3, 76.9, 67.7, 67.4, 66.8, 66.8, 51.8, 38.8, 30.8, 30.2, 27.2, 23.5; HRMS (ESI) *m/z* calcd. for $C_{30}H_{31}NO_8$ $(M+H)^+$ 533.2044, found 533.2050.

A.1.3. Synthesis and Characterization of Natural Products

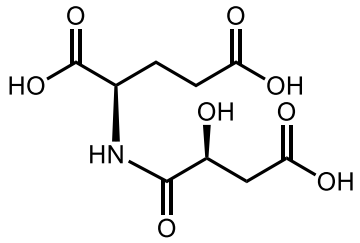
General Procedure for Hydrogenolysis (B): To a flame dried flask equipped with a magnetic stir bar were sequentially added fully benzylated precursor and Pd/C (10%, 5.0 mol%). The vial was sealed with a septum and pumped and back filled 3 times with nitrogen. This was followed by addition of dry ethanol (0.026 M) and careful addition of H_2 via balloon. The reaction mixture was stirred for 6 hours and then filtered through celite with DCM. Solvent was removed in vacuo without further purification.



L,L-malyl glutamate

(*S*)-2-((*S*)-3-Carboxy-2-hydroxypropanamido)pentanedioic acid (1)

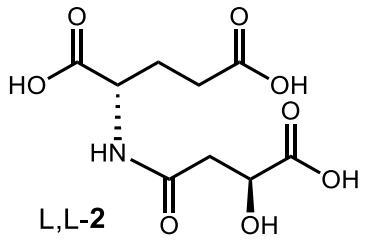
Prepared according to General Procedure B using L,L-10 (126 mg, 0.236 mmol) to afford LL-1 as a white solid (46 mg, 73% yield). IR (film) 3396, 2844, 1710, 1450, 1361 cm^{-1}; ^1H NMR (400 MHz, D_2O) δ 4.58 (dd, *J* = 4.0, 8.0 Hz, 1H), 4.47 (dd, *J* = 4.0, 8.0 Hz, 1H), 2.90 (dd, *J* = 4.0, 16 Hz, 1H), 2.79 (dd, *J* = 4.0, 16 Hz), 2.50 (t, *J* = 8.0 Hz, 2 H), 2.22-2.32 (m, 1H), 2.00-2.12, (m, 1H); ^{13}C NMR (100 MHz, D_2O) δ 177.1, 176.8, 174.6, 71.4, 66.9, 38.5, 28.7; HRMS (ESI) *m/z* calcd. for $C_9H_{14}NO_8$ $(M+H)^+$ 264.0714, found 264.0710.



L,D-malyl glutamate

(*R*)-2-((*S*)-3-Carboxy-2-hydroxypropanamido)pentanedioic acid (1)

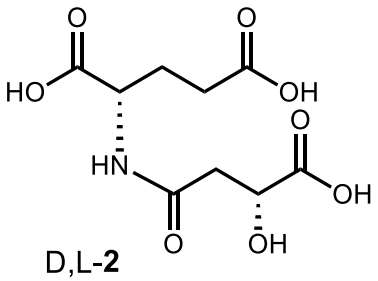
Prepared according to General Procedure B using L,L-10 (63 mg, 0.118 mmol) to afford LL-1 (22 mg, 71% yield). IR (film) 3396, 2844, 1710, 1450, 1361, cm^{-1}; ^{1}H NMR (400 MHz, D_2O) δ 4.57 (dd, *J* = 4.0, 8.0 Hz, 1H), 4.48 (dd, *J* = 4.0, 8.0 Hz, 1H), 2.90 (dd, *J* = 4.0, 16 Hz, 1H), 2.77 (dd, *J* = 4.0, 16 Hz), 2.50 (t, *J* = 8.0 Hz, 2 H), 2.22-2.32 (m, 1H), 2.04-2.12, (m, 1H); ^{13}C NMR (100 MHz, D_2O) δ 177.1, 176.7, 174.6, 133.7, 127.1, 71.4, 66.9, 38.6, 28.8; HRMS (ESI) *m/z* calcd. for $C_9H_{14}NO_8$ $(M+H)^+$ 264.0714, found 264.0719.



L,L-**2**

(*S*)-2-((*S*)-3-Carboxy-3-hydroxypropanamido)pentanedioic acid (2)

Prepared according to General Procedure B using **L,L-S1** (20 mg, 0.038 mmol) to afford **LL-1** (9 mg, 82% yield). IR (film) 3352, 3034, 1733, 1536, 1386, cm^{-1}; ^{1}H NMR (400 MHz, D_2O) δ 4.60 (dd, *J* = 4.5, 7.8 Hz, 1H), 4.46 (dd, *J* = 4.7, 8.2 Hz, 1H), 3.20-3.33 (m, 1H), 2.82 (dd, *J* = 4.5, 15.2, 1H), 2.70-2.77 (m, 1H), 2.35-2.55 (m, 3H), 2.15-2.34 (m, 1H), 1.90-2.05 (m, 1H); ^{13}C NMR (100 MHz, D_2O) δ 177.2, 175.3, 174.9, 174.5, 67.9, 66.9, 57.5, 51.9, 48.9, 38.6, 29.9, 16.8.; HRMS (ESI) *m/z* calcd for $C_9H_{14}NO_8$ $(M+H)^+$ 264.0714, found 264.0721. The specific rotation $[\alpha]^{24}_D$ = -13.6° (*c* 0.1, MeOH).



D,L-**2**

(*S*)-2-((*R*)-3-Carboxy-3-hydroxypropanamido)pentanedioic acid (2)

Prepared according to General Procedure B using D,L-S1 (10 mg, 0.019 mmol) to afford D,L-2 (3.8 mg, 78% yield). IR (film) 3350, 2853, 1730, 1260, 1170, cm^{-1}; ^{1}H NMR (400 MHz, D_2O) δ 4.59 (dd, *J* = 4.4, 7.8 Hz, 1H), 4.46 (dd, *J* = 5.1, 9.0 Hz, 1H), 2.83 (dd, *J* = 4.5, 15.2, 1H), 2.77 (dd, *J* = 7.9, 15.2, 1H), 2.51 (t, *J* = 7.5, 2H), 2.15-2.28; ^{13}C NMR (100 MHz, D_2O) δ 177.1, 175.3, 174.8, 174.5, 67.9, 57.4, 51.7, 38.5, 29.9, 25.6, 16.7; HRMS (ESI) *m/z* calcd for $C_9H_{14}NO_8$ $(M+H)^+$ 264.0714, found 264.0720.

A.1.4. ^{1}H and ^{13}C NMR Spectra of Synthesized Malylglutamate Isomers

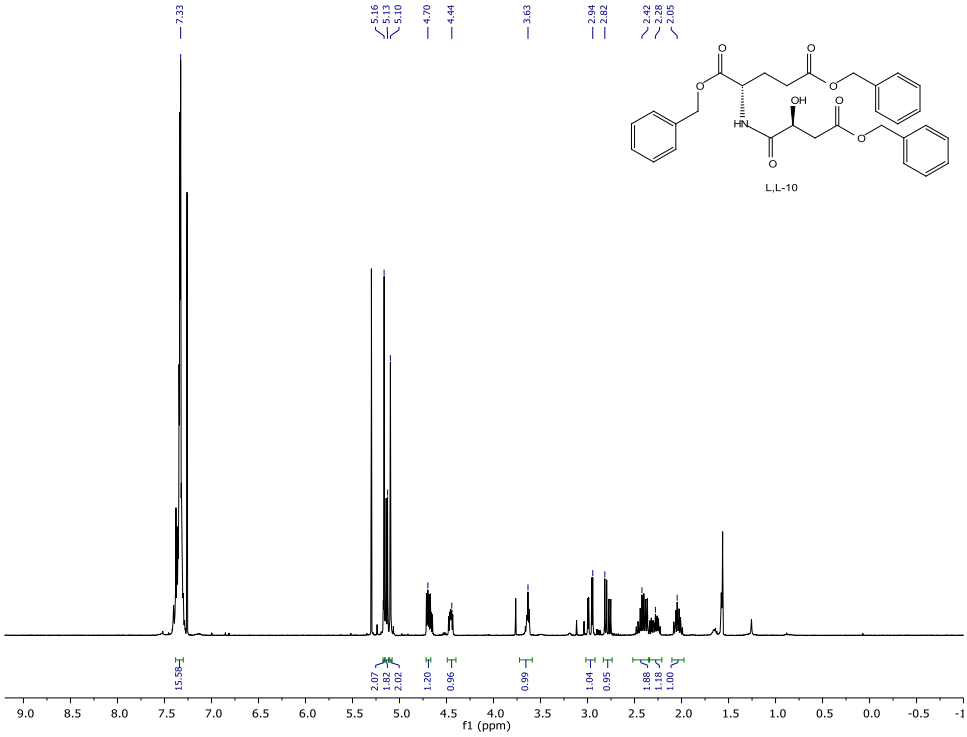


Figure A.3. ^{1}H NMR spectrum of benzyl ether protected L,L-α-malylglutamate.

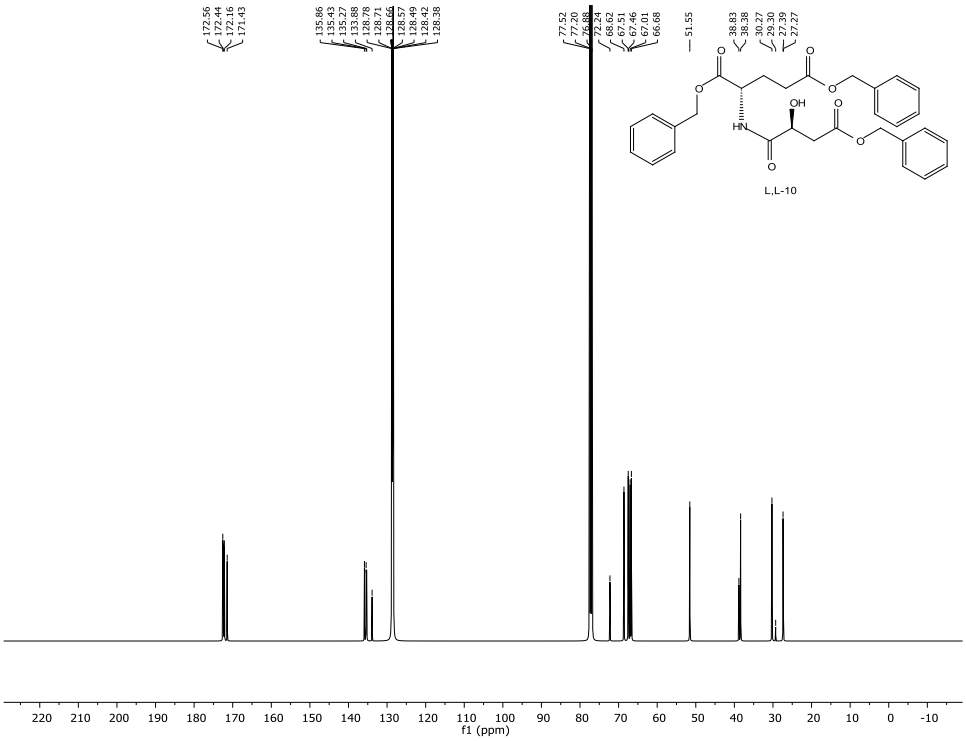


Figure A.4. ^{13}C NMR spectrum of benzyl ether protected L,L-α-malylglutamate.

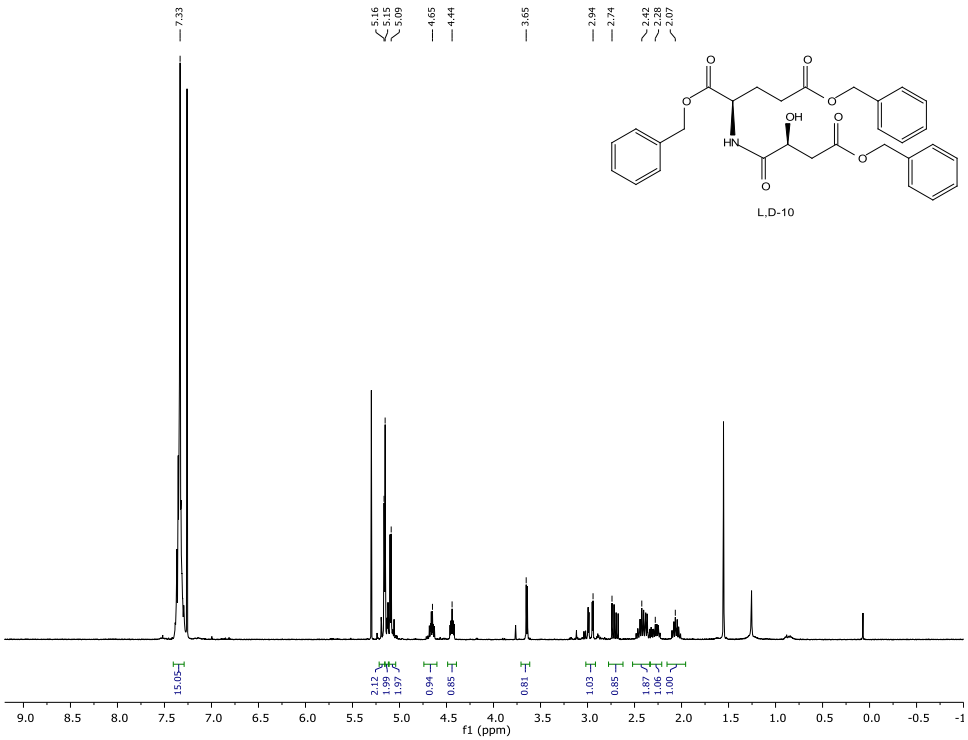


Figure A.5. ^{1}H NMR spectrum of benzyl ether protected L,D-α-malylglutamate.

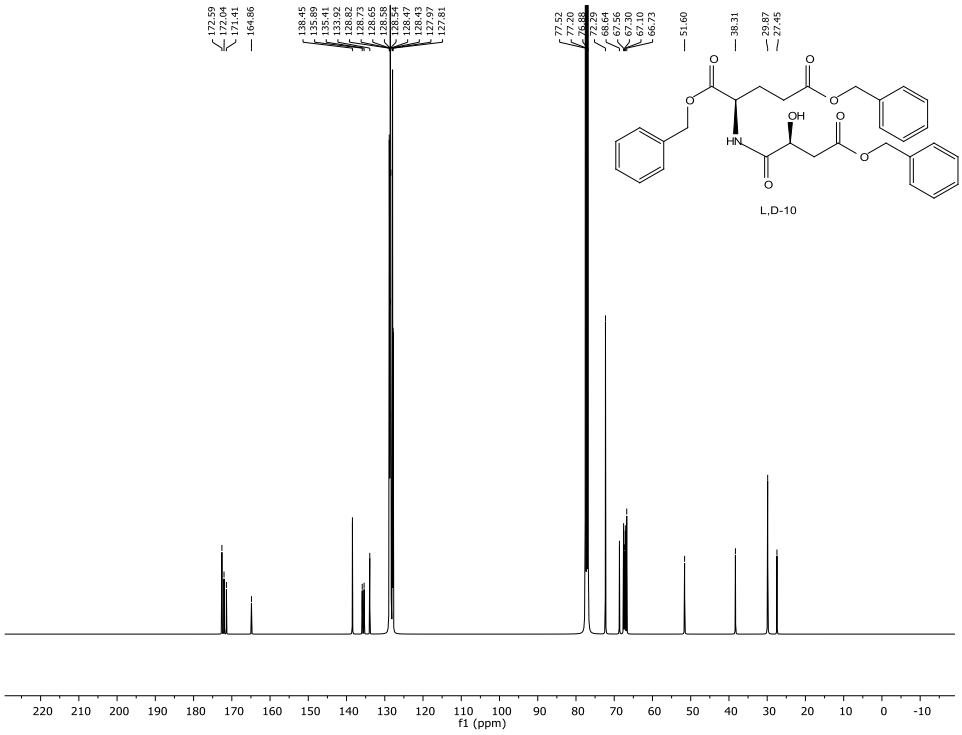


Figure A.6. ^{13}C NMR spectrum of benzyl ether protected L,D-α-malylglutamate.

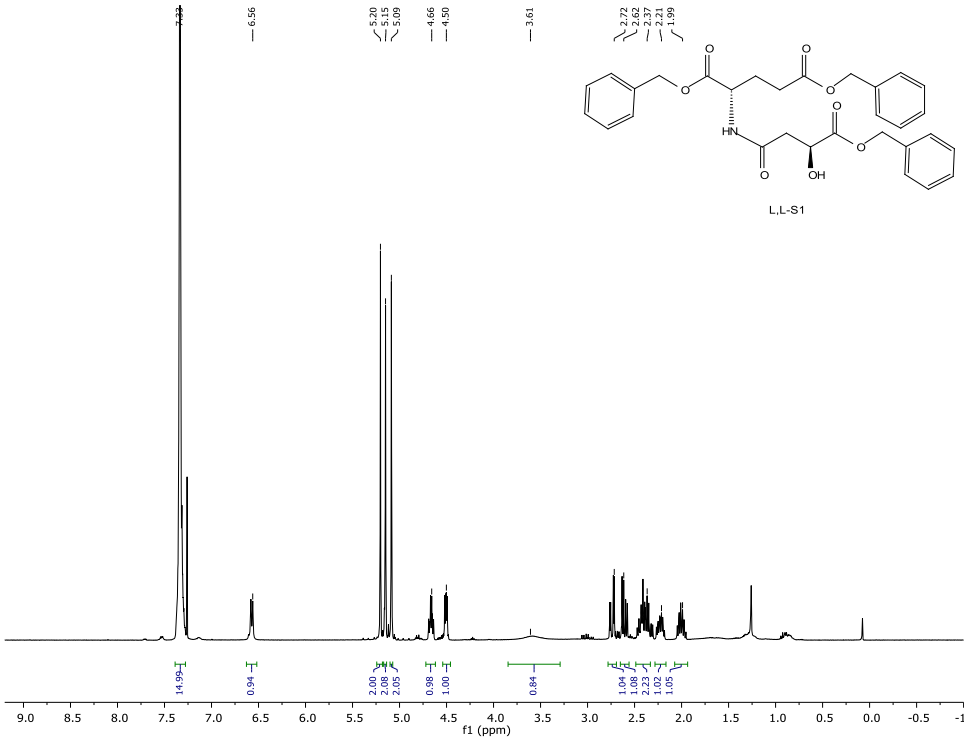


Figure A.7. ^{1}H NMR spectrum of benzyl ether protected L,L-β-malylglutamate.

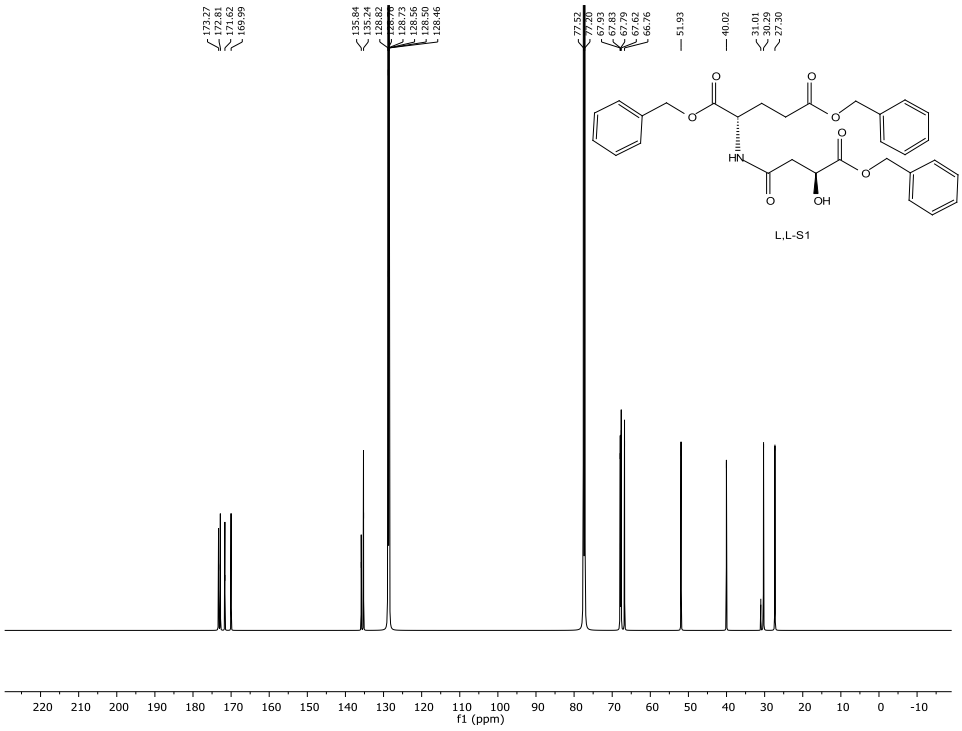


Figure A.8. ^{13}C NMR spectrum of benzyl ether protected L,L-β-malylglutamate.

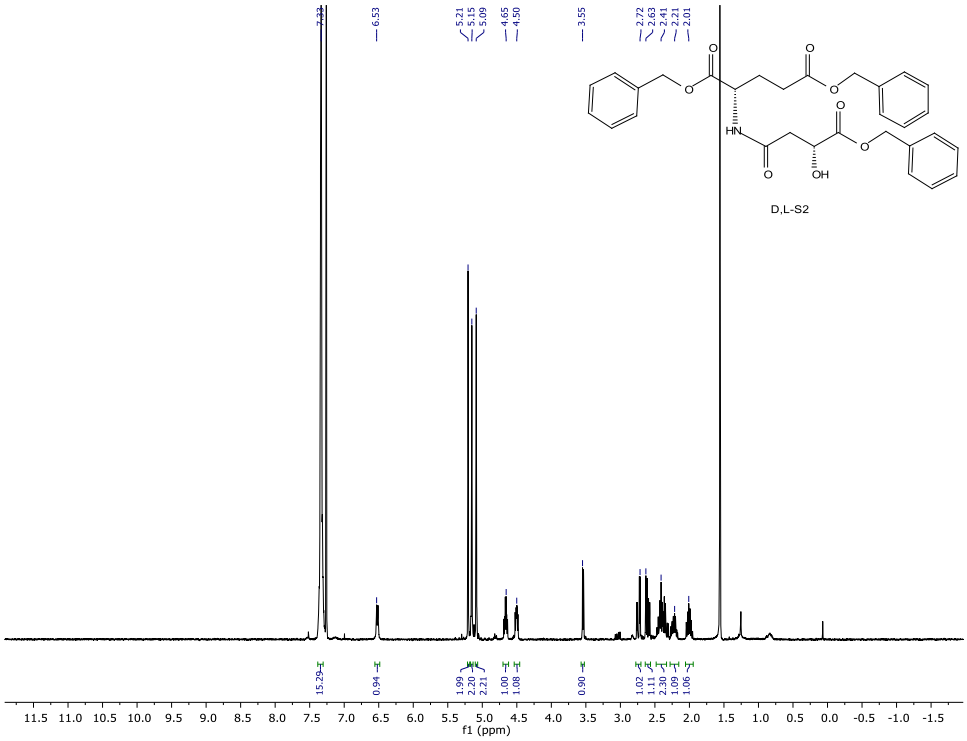


Figure A.9. ^{1}H NMR spectrum of benzyl ether protected D,L-β-malylglutamate.

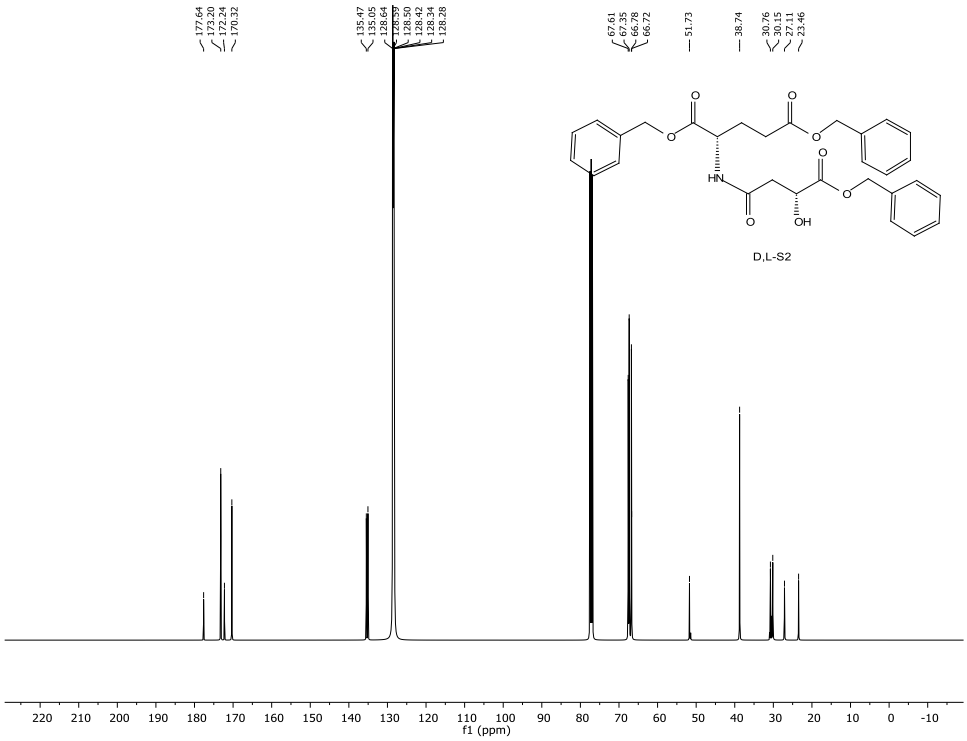


Figure A.10. ^{13}C NMR spectrum of benzyl ether protected D,L-β-malylglutamate.

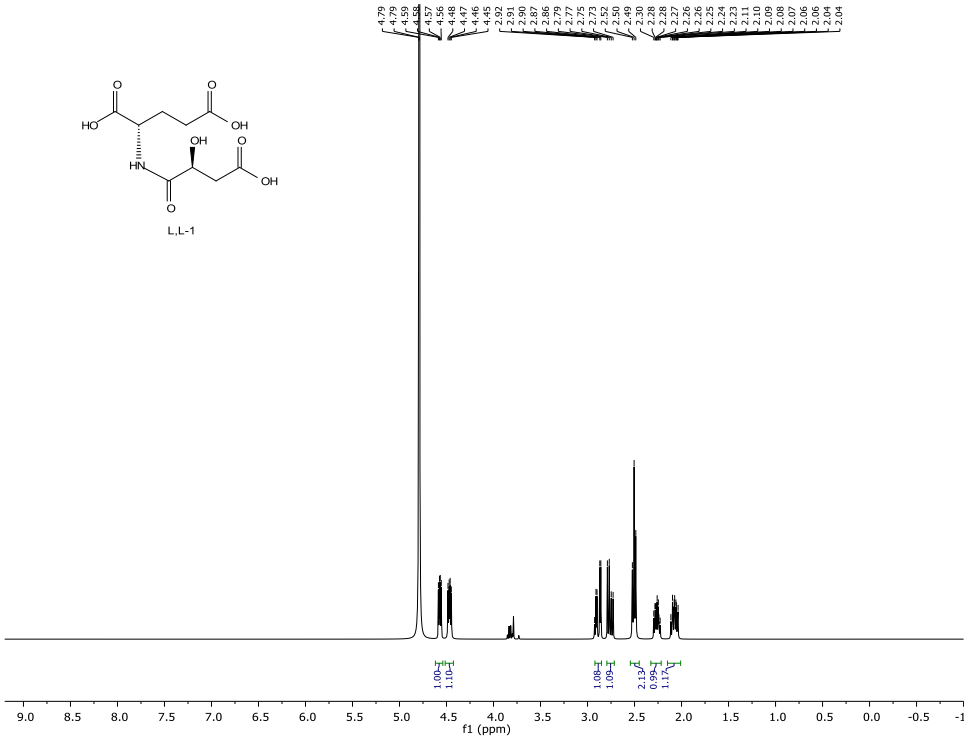


Figure A.11. ^{1}H NMR spectrum of L,L-α-malylglutamate.

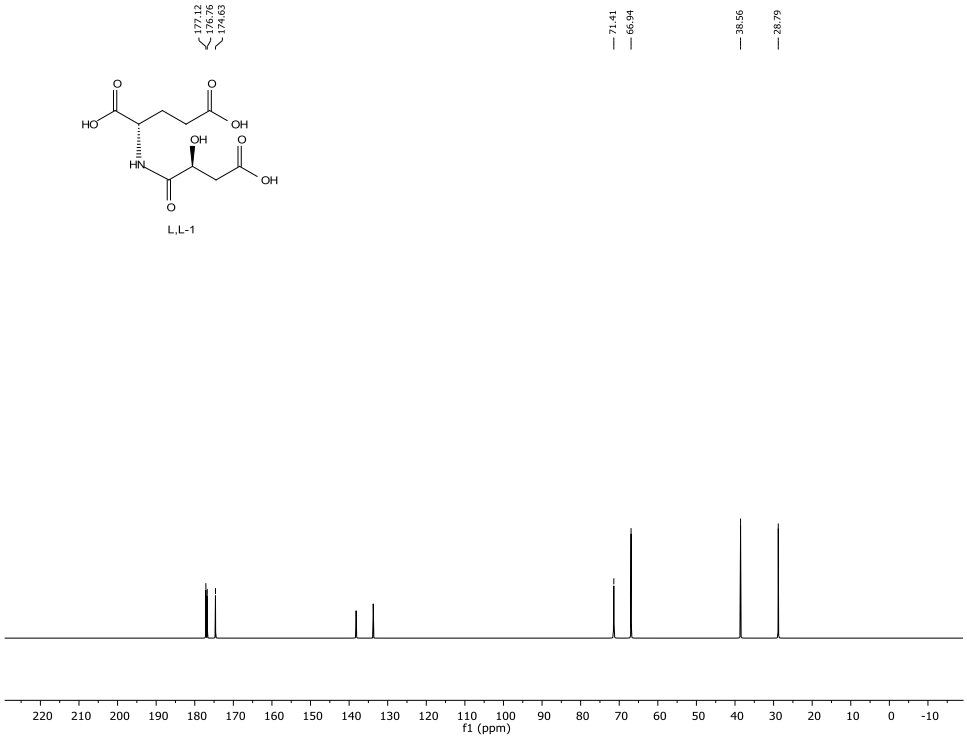


Figure A.12. ^{13}C NMR spectrum of L,L-α-malylglutamate.

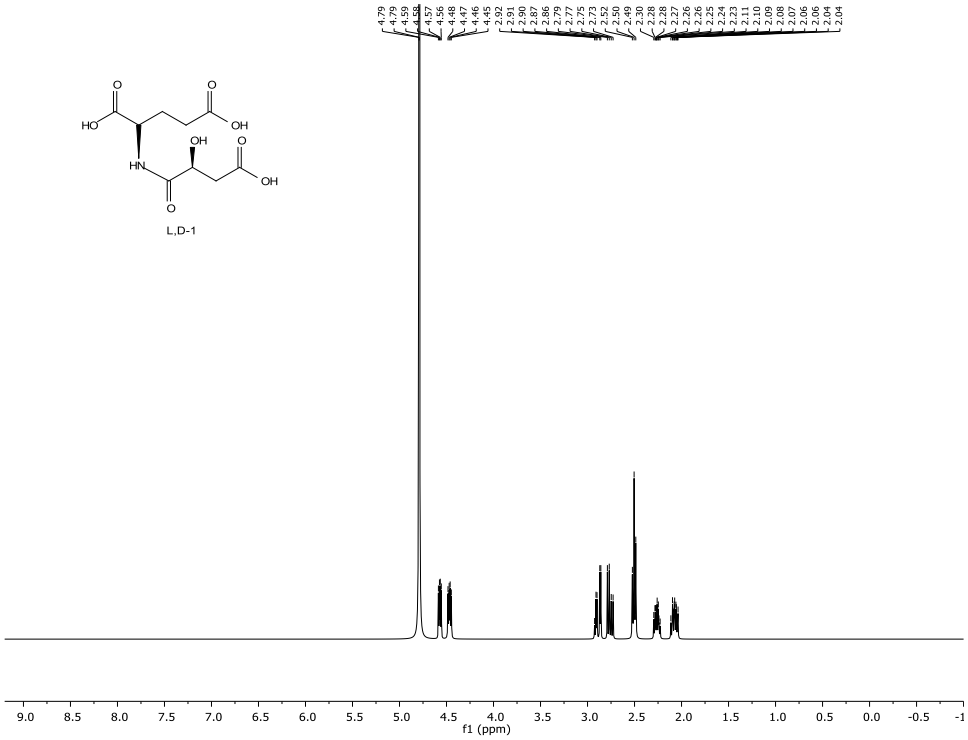


Figure A.13. ^{1}H NMR spectrum of L,D-α-malylglutamate.

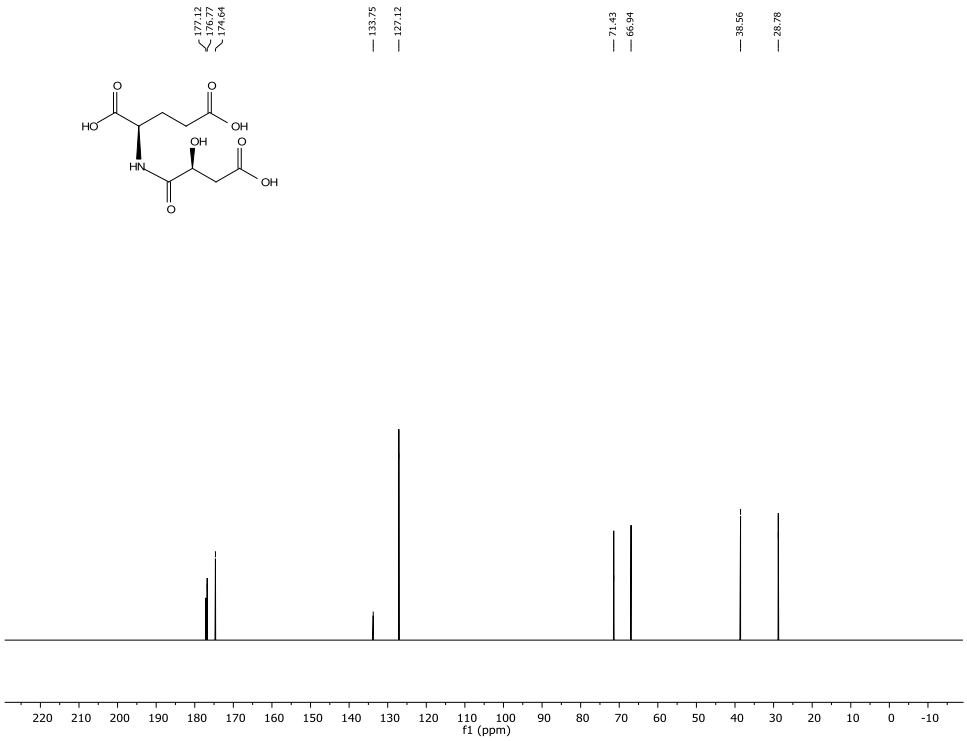


Figure A.14. ^{13}C NMR spectrum of L,D-α-malylglutamate.

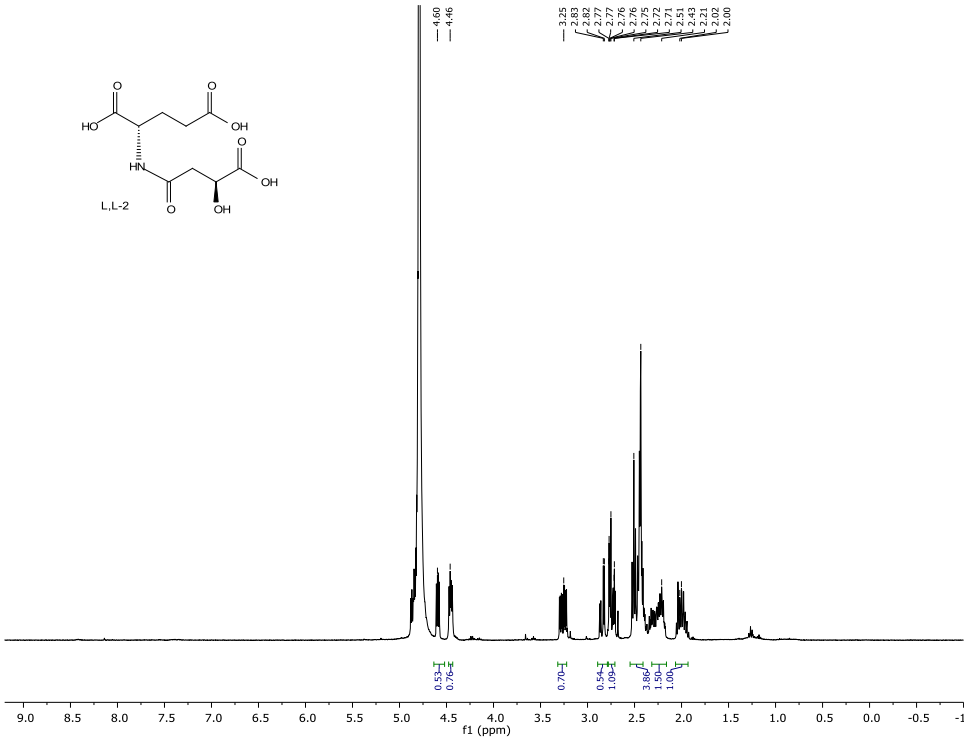


Figure A.15. ^{1}H NMR spectrum of L,L-*β*-malylglutamate.

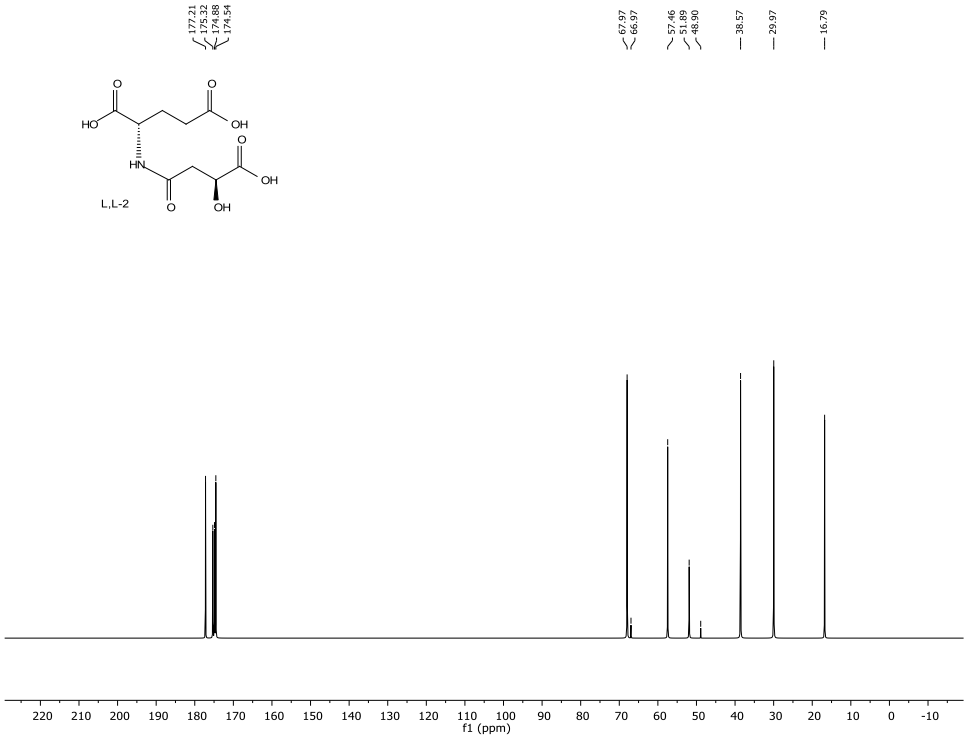


Figure A.16. ^{13}C NMR spectrum of L,L-β-malylglutamate.

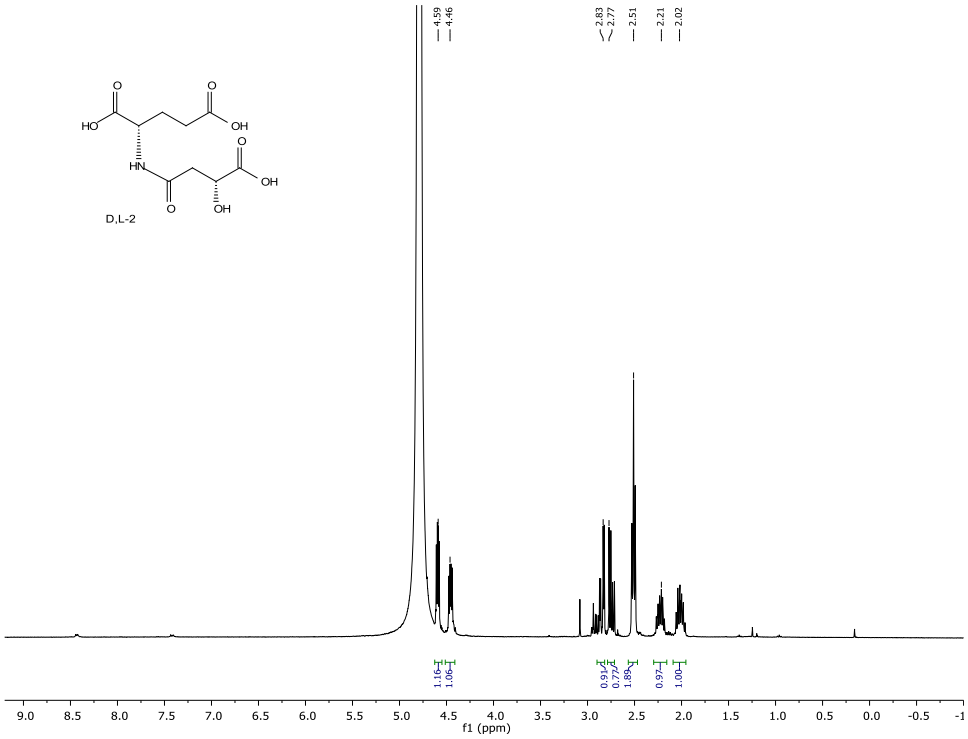


Figure A.17. ^{1}H NMR spectrum of D,L-β-malylglutamate.

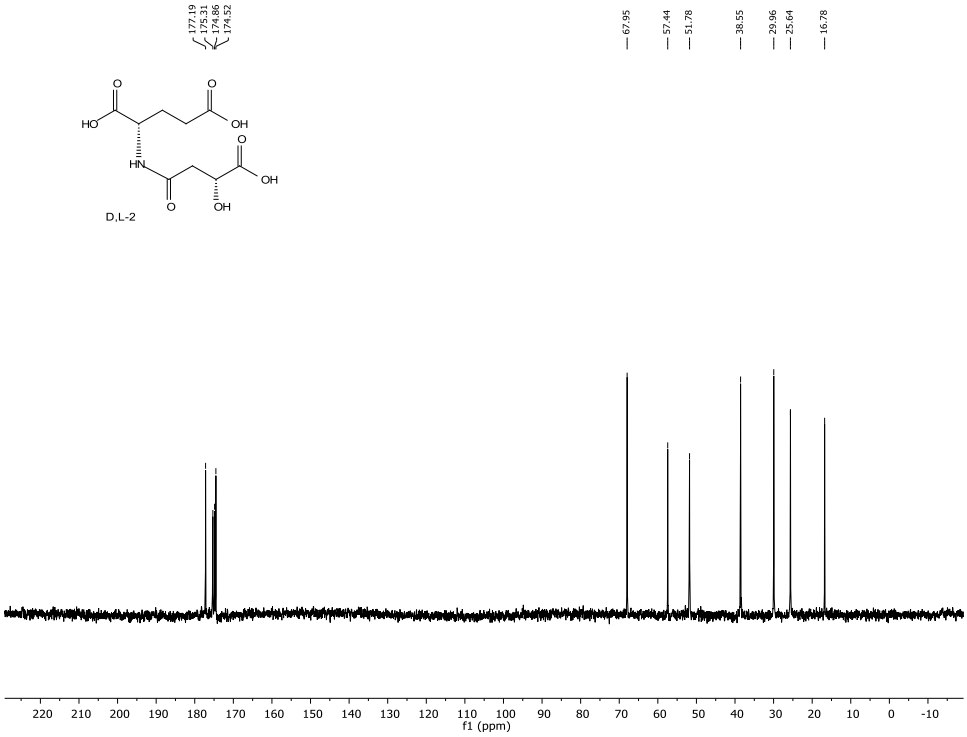


Figure A.18. ^{13}C NMR spectrum of D,L-β-malylglutamate.

A.2. References

1. Chan, T. M.; Kong, J.; McNamara, P.; Wong, J. K., Syntheses of potential degradation products of phenylephrine in otc products. *Synth. Commun.* **2008,** *38* (13), 2252-2260.

2. Du, Z.; Lu, Y.; Dai, X.; Zhang-Negrerie, D.; Gao, Q., The discovery of a facile access to the synthesis of nsaid dendritic prodrugs. *J. Chem. Res.* **2013,** *37* (3), 177-180.

3. Schobert, R.; Jagusch, C., An efficient synthesis of carlosic acid and other 5-carboxymethyltetronates from malates. *Synthesis* **2005,** *2005* (14), 2421-2425.

4. Dutton, F. E.; Lee, B. H., Epsilon-lactam analogs of the anthelmintic cyclodepsipeptide pf1022a. *Tetrahedron Lett.* **1998,** *39* (30), 5313-5316.